

Reliability and Validity of the Target Behaviour Severity Scale: A Pilot Study

Whitney Kerr. B. A. Hons.

Centre for Applied Disability Studies

Submitted in partial fulfillment
of the requirements for the degree of

Master of Arts

Faculty of Social Sciences, Brock University
St. Catharines, Ontario

© 2016

Abstract

This pilot study examined the reliability, validity, and, usability of the Target Behaviour Severity Scale (TBSS), which was designed for use as a program evaluation tool for ABA-based clinical services. The TBSS measures four dimensions of behaviour: frequency, intensity, duration, and discrimination, and yields a total severity score. Participants included eleven consultants and 25 caregivers of individuals with DD receiving publically funded behavioural services to treat dangerous behaviour, tantrums, disruptive behaviour, and skill deficits. The total severity score had acceptable internal consistency (consultant $\alpha = .714$, caregiver $\alpha = .730$). Examination of interrater reliability revealed that caregivers' total severity scores were significantly higher than consultants' scores though they shared a strong and significant positive relationship. The reliability coefficients of the dimension scales were variable, suggesting the need for some refinements. The results of this pilot study suggest that the TBSS holds promise as a program evaluation tool.

Keywords: developmental disabilities, intellectual disability, autism spectrum disorder, problem behaviour, program evaluation

Acknowledgements

To create a thesis is to create a body of work. As a visual thinker, I have identified each part of the body that all the people involved in this process represent. Thank you Dave for your precise feedback and attention to detail. Melissa, thank you for your vigilance and impartation of mastered APA guidelines. Those details have been the perfected hair to the thesis body. I am so very grateful for the vision development and input of my committee members. Thank you Dr. Feldman, Dr. Isaacs, and Dr. Yu, for all of the feedback and thoughts that you have put into this work. You symbolize the skin, pulling this all together, despite my scatter-brained tendencies. Rosemary, thank you for translating ABA from a concept into a reality for me. Despite my learning limitations, you have put forth every learning strategy possible to help optimize my performance. You have literally gone above and beyond supervision-providing clinical, professional, and personal advice, support, and direction, invariably changing my future. You represent the muscles of the body, providing substance to my arguments and reviving my work energy whenever needed. God, you are the veins that have given me renewed breath throughout this entire process. Steve, you are my absolute backbone. You have been the stability through every general and specific step of the thesis process (i.e., meetings, disagreements, conflicts, gains, computer crashes, missed phone calls, cancelled dates). Thank you so much for seeing me through this until the end. This research was supported by the Social Sciences and Humanities Research Council of Canada and the Dean of Graduate Studies Spring Research. I am very grateful for this financial support along with funding from the Dean of Graduate Studies for without it, this project would not have been completed and I would not be in the good health that I am in now. Thanks for being lungs! Together, we have worked to form one body, a thesis. I am beyond grateful for the additions and contributions of each member. Thank you.



Table of Contents

Acknowledgements	iii
List of Tables	vi
List of Figures	vii
Background.....	1
Research Objectives	3
Literature Review	3
Characteristics and Conditions of Developmental Disabilities	4
Problem Behaviour and DD	5
Impact of Problem Behaviour on Individuals and Caregivers	7
The Financial Impact of DD on Society	13
ABA as Treatment Model	14
ABA Service Delivery	19
Measuring Problem Behaviour	23
Program Evaluation	24
The Need for Systematic Measurement of Problem Behaviour	26
Existing Measures of Problem Behaviour	27
Gaps in Existing Measures of Problem Behaviour	30
The Need for Measures that are Socially Valid	31
Rationale	34
Purpose	34
Research Questions	35
Method	35
Recruitment	35
Participants	37
Measures	38
Procedure	41
Results	42
Final Sample	42
Target Behaviours Rated on the TBSS	43
Consultant's and Caregiver's Responses on the TBSS	44
Internal Consistency	48
Interrater Reliability	50
Face Validity	53
Usability	53
Consultant and Caregiver Feedback on TBSS	55
Discussion.....	56
Research Question 1: Does the TBSS have acceptable internal consistency?.....	57
Research Question 2: What is the inter-rater reliability between consultants and caregivers on the TBSS?	58
Research Question 3: What are consultant's and caregiver's opinions of the face validity and usability of the TBSS?	60
Limitations	61
Future Research	62
Conclusion.....	63
References.....	65

List of Tables

Table 1. Target Behaviours Rated on the TBSS	44
Table 2. Summary of Consultant's and Caregivers Response's to the TBSS	45
Table 3. Internal Consistency of the TBSS Dimension Totals	49
Table 4. Consultant's and Caregiver's TBSS Ratings	51
Table 5. Cohen's 'Rule of Thumb' for Correlation Coefficient Interpretation	52
Table 6. Correlations Between Consultant's and Caregiver's Responses	52
Table 7. Correlations Between Total Dimension of TBSS from Consultants	53
Table 8. Correlations Between Total Dimensions of TBSS from Caregivers	53
Table 9. Hard to Understand Questions for Consultants and Caregivers	54

List of Figures

Figure 1. Breakdown of TBSS Questions Across Dimensions of Problem Behaviour.....	38
Figure 2. Supplementary Feedback Questions From the TBSS	41

Reliability and Validity of the Target Behaviour Severity Scale (TBSS)

Background

Research has consistently shown that individuals with developmental disabilities (DD) such as intellectual disabilities (ID) and autism spectrum disorders (ASD), display higher rates of problem behaviour than the general population (Baker et al., 2003; Cooper, Smiley, Morrison, Williamson, & Allan, 2007; Hill & Furniss, 2006; Kurtz et al., 2003; Lancioni, Singh, O'Reilly, & Sigafoos, 2009; McClintock, Hall, & Oliver, 2003). These problem behaviours can negatively impact the individuals who engage in problem behaviours and can also have a significant impact on the lives of their caregivers (Brown, McIntyre, Crnic, Baker, & Blacher, 2011; McIntyre, Blacher, & Baker, 2002; Renty & Roeyers, 2006). Problem behaviour accounts for a significant portion of decreased parent well-being (Baker, Blacher, Crnic, & Edelbrock, 2002; Baker et al., 2003) and the reduction of problem behaviour can increase quality of life for the individual with DD as well as their caregivers (Feldman, Condillac, Tough, Hunt, & Griffiths, 2002).

Applied Behaviour Analysis (ABA) is an evidenced-based practice that relies on the scientific method to decrease problem behaviour and increase skills. Research demonstrating the efficacy of behavioural procedures for treating challenging behaviours in controlled settings are plentiful (e.g., Carr et al., 1999; Horner, Carr, Strain, Todd, & Reed, 2002; LaVigna & Willis, 2012; Sugai et al., 2000) but there are fewer studies examining the effectiveness of specific ABA procedures implemented by caregivers in real world settings (e.g., Dib & Sturmey, 2007; Doggett, Edwards, Moore, Tingstrom, & Wilczynski, 2001; Feldman et al., 2002). In Ontario, the majority of publicly funded behavioural services available to treat problem behaviour in individuals with DD employ a mediator model. In this model, trained consultants complete behavioural assessments and design interventions to be implemented by caregivers. Outcome

data reporting the effectiveness of these programs in reducing problem behaviour and increasing skills are not typically available on a program-wide basis. This is likely related to the fact that ABA interventions are based on the results of an individualized functional assessment (Steege, Mace, Perry, & Longenecker, 2007) and ongoing examination of data. This individualized approach increases effectiveness of service. However, it can pose challenges for carrying out broad-based program evaluation, which requires more consistent measurement across various behaviours, mediators, treatments, and environments (National Research Council [NRC], 2011; Weitlauf et al., 2014, p. ES-7, ES-18). Given the cost of delivering these services in Ontario alone, stakeholders require program evaluation to determine outcomes both within and across settings and agencies that provide ABA-based services. Program evaluation can assist in determining if the investment in ABA services is leading to the desired outcome, therefore validating or invalidating the efficacy investment decisions (Posovac & Carey, 2007).

This thesis is focused on the development of the TBSS as a measure of the severity of problem behaviour to be used within a program evaluation system specifically designed to evaluate ABA programs (Condillac, 2009a). There are three parts to that program evaluation system, (a) a tracking system to record key aspects of client characteristics, assessment procedures and outcomes, treatment procedures and outcomes, and follow up status (Condillac, 2009c); (b) a measure of the impact that the problem behaviour has on life quality for the individual who is the target of intervention, and the individuals in their natural environments (Condillac, 2009b); and (c) the Target Behaviour Severity Scale (TBSS), a measure of the severity of problem behaviour that can be completed by clinicians and by the natural caregivers (Condillac, 2009d). The present study will focus on a field test of the TBSS as a measure of the

severity of the problem behaviour exhibited by 27 individuals with DD who are the recipients of ABA-based services.

Research Objectives

The general objectives of the current study are to examine the reliability, validity, and usability of the TBSS when used by consultants and caregivers. Stolarova, Wolf, Rinker, and Brielmann (2014), described reliability as the precision of a measure and “its capacity to produce constant, similar results” (p. 1). Reliability in this study will be examined through calculations of internal consistency and inter-rater reliability. Validity refers to the accuracy of a test to measure the concept or construct that it is designed to measure. Usability refers to the reported ease of use by different raters. The three research objectives of the current study are:

1. To determine the internal consistency of the TBSS.
2. To determine the inter-rater reliability between consultants and caregivers on the TBSS.
3. To gather consultant’s and caregiver’s opinions of the face validity and usability of the TBSS.

This study is an important step in refining this measure and contributing to the development of a much needed program evaluation system for behaviour analytic services in Ontario.

Literature Review

The following section provides pertinent literature relating to the need for and development of the TBSS for use by behaviour consultants, caregivers, and other relevant stakeholders to provide a common measure of the severity of problem behaviour over the course of assessment and intervention. First, the characteristics and conditions of individuals with DD

will be described and the effects of problem behaviour on the lives of individuals with DD and their caregivers will be discussed. Second, an overview of ABA methodologies including assessments, intervention, and observation and monitoring will be provided. Third, ABA services used in Ontario will be described, emphasizing some of the challenges that exist within the model. Fourth, the impact that these challenges have on program evaluation will be discussed, and the need for a measure of severity will be examined. Fifth, existing measures of problem behaviours will be reviewed and their limitations for program evaluation purposes will be presented. Finally, the dearth of measures to assess the severity of a range of specific problem behaviours will be discussed and a new measure, the TBSS will be described.

Characteristics and Conditions of Developmental Disabilities

A developmental disability (DD) is a cognitive and adaptive impairment that impacts learning, language, and behaviour (Services and Supports to Promote the Social Inclusion of Persons with Developmental Disabilities Act; SIPDDA, 2008) and has long-term effects on the individual (American Psychiatric Association [APA], 2013). Some examples of DD include cerebral palsy, Down's syndrome, intellectual disability, and autism (Centers for Disease Control and Prevention [CDC], 2010). Throughout this thesis, 'DD' will be used to describe autism spectrum disorders and/or intellectual disabilities.

Intellectual disabilities (ID) are characterized by impairments in intellectual functioning (e.g., IQ level lower than 70) and maladaptive functioning in conceptual, social, or practical domains (APA, 2013). The Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5) states that specific areas of intellectual functioning such as problem solving, executive functioning, and learning may be affected. The diagnostic criteria for ID was recently modified to accommodate a movement away from a deficit-based model of mental retardation to a

“functional classification” (Harris, 2006) relating to support needs (Brown, Ouellette-Kuntz, Bielska, & Elliott, 2009).

Autism spectrum disorders (ASD) are a neurodevelopmental disorder characterized by behavioural excesses and deficits in relation to social communication as well as restricted and repetitive interests and stereotypic behaviour (APA, 2013). Individuals with ASD have social communication difficulties such as impairments in social-emotional reciprocity, nonverbal communication, or difficulty in forming or maintaining relationships (CDC, 2010). Individuals with ASD have restricted and repetitive interests, such as rigid play behaviour, strict adherence to routines, or an insistence on sameness (Weitlauf et al., 2014). Vocal stereotypy, hand flapping, or echolalia, are some examples of stereotypic behaviour (Schreibman, Heyser, & Stahmer, 1999). Since the publication of the DSM- 5 in 2013, the three deficits of ASD included in the fourth edition of the DSM (DSM-IV) were divided into two domains: impairments in social-communication and restricted and repetitive behaviours in order to reduce misunderstandings between these two areas (Lauritsen, 2013). Comorbidity rates of ASD and ID under the text revised version of the DSM-IV (DSM-IV-TR) showed that up to 78.7% of individuals with ASD also had an ID (Mannion, Leader, & Healy, 2013). Although modifications have been made to ASD under DSM-5 criteria, current prevalence data includes individuals based on DSM-IV-TR (Baio, 2014), possibly rendering an inaccurate representation of the actual prevalence (Maenner et al., 2014). Nevertheless, the most current prevalence rate of children with ASD who have a comorbid ID diagnosis is 31% (Baio, 2014).

Problem Behaviour and DD

In the literature, problem behaviour has also been referred to as challenging, maladaptive, or aberrant behaviour. In this thesis, the term ‘problem behaviour’ will be used because not all

problem behaviour is “challenging,” or “aberrant”. In an operant understanding, behaviour is functionally related to the contingencies that maintain it and therefore can be quite “adaptive” (Carr & Durand, 1985; Iwata, Dorsey, Slifer, Bauman, & Richman, 1994). Some of the most prevalent problem behaviours include aggression, self-injury, and stereotypic behaviours (Rojahn, Matson, Lott, Esbensen, & Smalls, 2001).

As previously noted, problem behaviour is more common among individuals with DD than those in the general population (Baker et al., 2003; Cooper, Smiley, Morrison, Williamson, & Allan, 2007; McClintock et al., 2003). Specific prevalence rates of problem behaviour have differed across reports due to “differences in operational definitions, ASD diagnosis, measurement instrument, and sample characteristics” (Didden et al., 2012, p. 41). However, some reports estimate that nearly 50% of children with DDs (Lecavalier, 2006) and between 40-60% of adults with ID exhibit problem behaviours (Kats, Payne, Parlier, & Priven, 2013; McCarthy et al., 2010).

Stereotypic behaviour, tantrums, destruction, aggression, and self-injurious behaviour have been cited as some of the most prevalent problem behaviour exhibited by children with DD (Baker et al., 2003; Didden et al., 2012). There are some topographical differences of problem behaviour between individuals with ASD and individuals with ID (Didden et al., 2012). Prevalence rates increase twofold for individuals with a dual diagnosis of ASD and ID compared to individuals with ID alone (Kats et al., 2013). Matson and Rivet (2008) compared prevalence of problem behaviour in individuals with ASD and DD and found that problem behaviour was exhibited by 28% of individuals with autistic disorder, 11% of individuals with pervasive developmental disorder not otherwise, and only 6% individuals with ID. Not only do the rates of problem behaviours differ across populations, but the topography of the problem also varies.

McClintock and colleagues (2003) evaluated these differences in a meta-analysis. They found that individuals with ASD most often engaged in aggression, stereotypy, and disruptive behaviour, while individuals with ID most often engaged in self-injury or stereotypy. Aggression was associated with communication deficits for both groups and was more prevalent among males. This finding is common in the literature; problem behaviour is more common among males than females (Eisenhower, Baker, & Blacher, 2005). A likely explanation for the higher prevalence of problem behaviour in individuals with ASD could be the related communication deficits. Koegel and Koegel (2006), and Durand & Carr (1991) identified that the lack of verbal communication can often lead to problem behaviour if the individual's needs cannot be understood or met.

Individuals with DD display problem behaviour that can impact aspects of their own lives as well as the lives of their families (Lach et al., 2009; Renty & Roeyers, 2006; McIntyre et al., 2002). Furthermore, individuals who display problem behaviour often require additional services, which increase the cost of meeting their needs (Knapp, Comas-Herrera, Astin, Beecham, & Pendaries, 2005). For example, ABA-based services have been used to reduce problem behaviours as well as increase skills (Heyvaert, Saenen, Campbell, Maes, & Onghena, 2014; Horner et al., 2002; LaVigna & Willis, 2012; Sugai et al., 2000). Despite their potential for long-term benefit, these services contribute to the cost of supporting individuals with DD (Ganz, 2007). Program evaluation can be used to determine if ABA services lead to desired outcomes such as reductions in problem behaviour or increases in adaptive skills.

Impact of Problem Behaviour on Individuals and Caregivers

The severity of problem behaviour can also be measured by the effects it has on the individual's life or the life of their caregiver (Didden et al., 2012). For individuals with DD as

well as their caregivers, problem behaviour can have a tremendous negative impact on the life quality and health (Donenberg & Baker, 1993; Lach et al., 2009; McIntyre et al., 2002). The following is a review of the effects of problem behaviour on individuals with DD, their caregivers, and the interactions that they share.

Effects of problem behaviour on the individual. Problem behaviour impacts the individual who engages in it. Problem behaviour can have adverse effects on the individual's mental health (Emerson and Hatton, 2007), social acceptance (Didden et al., 2012; Rojahn et al., 2001), and can decrease learning opportunities (Ahearn, Clark, MacDonald, & Chung, 2007; Giles et al., 2012; Lanovaz & Sladeczek, 2012).

Mental health. Problem behaviour is associated with mental health challenges. The prevalence of mental health challenges among individuals with ID has been difficult to determine due to a myriad of challenges including biased sampling, measurement, and specificity in presentation of prevalence rates (for additional examples, see Cooper et al., 2007). Despite these difficulties, a study by Cooper and colleagues (2007) indicated that more than one third of the sample had a mental health issue, a rate that has been fairly consistent with other studies (Buckles, Luckasson, & Keefe, 2013). High rates of comorbid psychiatric and conduct disorders are common among children with DD (Einfeld, Ellis, & Emerson, 2011; Emerson & Hatton, 2007). Children with disabilities are at greater risk of developing mental health challenges than their typically developing peers (Cadman, Boyle, Szatmari, & Offord, 1987). If untreated, problem behaviour can persist into adulthood (Matson & Rivet, 2008). Reiss and Rojahn (1993) found a significant relationship between children with DD and mental health challenges into adulthood. Moss, Prosser, Costello, and Simpson (1998) demonstrated that as the severity of problem behaviour increased, the likeliness of depression increased fourfold.

Social acceptance. Problem behaviour is associated with social exclusion (Sigafoos, Arthur, & O'Reilly, 2003). Individuals with DD may experience low self-esteem (Patchin & Hinduja, 2010) due to social exclusion in academic (Cooney, Jahoda, Gumley, & Knott, 2006), community (Gray 1993), and online environments (Didden et al., 2009). Sigafoos et al. (2003) found that even further isolation from community involvement occurred when the individual exhibited problem behaviour.

Academic. Problem behaviours regularly impede learning and potential social opportunities within the classroom (Ahearn et al., 2007; Macdonald et al., 2007). Problem behaviour can directly impact adaptive functioning, influencing the child's classroom placement (De Bildt, Sytema, Kraijer, Sparrow, & Minderaa, 2005). Often, behaviour management can take precedence over education, thereby limiting the overall opportunities to learn through academic instruction (Alberta Learning, 2003).

Employment. If untreated, problem behaviours are likely to continue into adulthood, impacting employment opportunities (Didden et al., 2012; Matson & Rivet, 2008). Aggression, self-injury, and property destruction are risk factors for employers who are not able to support the individual's needs, nor cope with behavioural challenges (Hendricks, 2010). Syndrome specific behaviours have been reported to interfere with work relationships and expectations (Smith & Philippen, 2005), lead to social exclusion and stigmatization (Winn & Hay, 2009), and have a negative impact on self-esteem (Patchin & Hinduja, 2010). Few behavioural problems as well as self-determination, and being high functioning were found to be key predictors of successful employment for individuals with ID (Martorell, Gutierrez-Recacha, Pereda, & Ayuso-Mateos, 2008). Although job support has been found to improve employment success (Martorell et al., 2008) it may not be easily accessible (McDonough & Revell, 2010).

Effects of problem behaviour on caregivers. Problem behaviour influences caregiver mental health (Benson & Karloff, 2009; Lach et al., 2009; Abbeduto et al., 2004; McIntyre et al., 2002), marriage (Hartley et al., 2010) and can lead to additional costs for families (Knapp et al., 2005). Moreover, improved problem behaviour is also associated with improved life quality for individuals with DD and their families (Feldman et al., 2002). It is important to note that having a child with DD can have both positive and negative impacts on parents and families (Perry, 2004; Blacher & Baker, 2007; Minnes, Perry, & Weiss, 2015). In fact, parents of children with DDs share similar positive and negative experiences as parents of children without disabilities (Blacher & Baker, 2007). They also found that parental recognition of the positive impact of their child with DD on their lives mitigated the typically negative impact of that child engaging in problem behaviour. As Baker et al. (2002) noted, when parents rated their own parenting stressors, differences were attributed to the presence or absence of problem behaviour, not to their child's disability. The negative impact of problem behaviours on these key indicators will be reviewed.

Mental health. Research has found that caregivers' mental health can be impacted when their child with DD exhibits problem behaviour, has emotional problems (Herring et al., 2006), or has mental health challenges (McIntyre et al., 2002). Lach and colleagues (2009) compared the differences of parents of typically developing children, those with children with neurodevelopmental delays, and parents of children who had both neurodevelopmental delays and problem behaviours. They found that not only did parents of children with neurodevelopmental disorders experience greater negative physical and psychological impacts, but the families of children who had children with both neurodevelopmental disorders and problem behaviours displayed between 2 and 3 times more negative health conditions than the

other groups. The effects of having a child with a disability who subsequently engage in problem behaviour are additive to parent health and mental health (Lach et al., 2009).

Problem behaviour plays a significant role in caregiver stress levels and can often lead to depression (Moes, 1995). Depression rates are higher among parents of children with ASD than other disabilities (Benson & Karloff, 2009; Abbeduto et al., 2004). Child behaviour problem, escape-avoidance parent coping, and low social support have been associated with elevated depression scores for caregivers of children with or at risk for DD (Feldman et al., 2007). Olsson and Hwang (2001) compared depression rates among parents of children with and without DD. They found that single and married mothers of children with ASD had the highest depression scores followed by mothers of children with ID. These scores were higher still than mothers of children without disabilities. Depression scores across fathers did not differ across groups. A common trend throughout the disability literature is that mothers experience greater impacts of having a child with a disability than fathers (Azeem et al., 2013; Hastings & Brown, 2002; Herring et al., 2006), possibly because mothers are the predominant caregivers (Statistics Canada, 2013). One Australian study noted high prevalence of anxiety and depression among parents of children with ASD with over 80% varying levels of anxiety and 60% experiencing some form of depression (Bitsika, & Sharpley, 2004). Similarly, parents of children with ID also experience higher levels and anxiety and depression than parents of typically developing children (Azeem et al., 2013). Azeem et al. (2013) demonstrated that mothers experienced 12% more symptoms of anxiety, depression, or both than fathers of children with ID.

Marriage. Having a child with a disability can add to marital stress and has been found to be associated with higher divorce rates than the typical population (Hartley et al., 2010). It has also been suggested that problem behaviour is an additional parent stressor (Walsh, Mulder, &

Tudor, 2013). Problem behaviour may be a precipitating variable of marital discord (Hartley et al., 2010; Lecavalier, Leone, & Wiltz, 2006) and may eventually lead to divorce (Hartley et al., 2010). Reports of an 80% divorce rate among parents of children with disabilities have not been empirically validated (Freedman, Kalb, Zablotzky, & Stuart, 2011). However, one peer-reviewed study by Hartley et al. (2010) found that divorce rates were almost 2 times higher (23.5%) for spouses who have a child with ASD, compared to 13.8% of parents of children without disabilities. Parents of children with ID experience less stress compared to spouses of a child with ASD (Griffith, Hastings, Nash, & Hill, 2010). Depending on the parental role, parents may experience different challenges with their child. A recent study found that mothers are more affected by behavioural challenges associated with ASD whereas fathers experience elevated stress levels in relation to their child's social acceptance (Shtayermman, 2013). Parents were particularly vulnerable to divorce as their children with ASD approached adolescence and adulthood (Hartley et al., 2010). Common existing marital difficulties such as communication challenges and lack of appreciation for one another (Mead, 2002) may compound additional stressors and challenges brought on by having a child with ASD (Benson, 2006). A common parent goal in treatment is to reduce problem behaviour, because it will lead to an improved life quality (Feldman et al., 2002).

Additional resources. Untreated problem behaviour is a high predictor of out of home placement when the family is unable to provide support the individual with DD (McIntyre et al., 2002). The decision to place an individual into fulltime care or community activities comes at a substantial cost to the caregivers (Cidav, Marucs, & Mandell, 2012; Horlin, Falkmer, Parsons, & Falkmer, 2014) because of the resources that are required to support the problem behaviour (Knapp et al., 2005) as well as a lack of social support available to caregivers in order to manage

the child's problem behaviour (Feldman et al., 2007). A large percentage of hospital day programs run in the United Kingdom required additional resources to support problem behaviour (Knapp et al., 2005). Lowe, Felce, and Blackman (1995) found that individuals who displayed externalizing problem behaviour such as aggression, property destruction, or social disruption were more likely to receive intervention whereas behaviours that were internalizing, for instance social withdrawal, were less likely to be referred to specialist teams and therefore further negatively impacted the quality of life of those individuals. Individuals who engage in problem behaviour are more vulnerable to lower standards of care, neglect, and even abuse (McGill, Emerson & Mansell, 1994).

In summary, while research has demonstrated that having a child with DD or ASD can have both a positive and a negative impact on families, the presence of problem behaviour is a predictor of negative impact (Baker et al., 2002), though this impact can be mitigated recognition of the positive impacts of having a child with DD (Blacher & Baker, 2007). The negative impacts of child problem behaviour on parental health, mental health, family relations, finances and general quality of life, suggest that the investment in programs and services to treat problem behaviour could have tremendous impact for the individual and their families.

The Financial Impact of DD on Society

After half a century of since Kanner's (1943) original description of ASD and its impact on individuals and families, the financial burden of ASD on families and society is beginning to be acknowledged. Support for individuals with ASD and ID have been divided in to indirect and direct costs translating to a total of \$35 billion annually in the United States each year (Ganz, 2006). Both direct and indirect costs are incurred for families, of which the greatest cost is therapy and caregiver lost productivity (Ganz, 2007). Indirect costs such as "missed time at

work, reduced work hours, switching to a lower-paying but more flexible job, or leaving the workforce” (Ganz, 2007, p. 344) are often incurred (Cidav et al., 2012). The financial impact increases for individuals with comorbid diagnoses such as ASD and ID (Buescher, Cidav, Knapp, & Mandell, 2014; Knapp et al., 2005). Problem behaviour is associated with even further costs for individuals with ID (Knapp et al., 2005) and individuals with ASD (Tureck, Matson, & Beighley, 2013). Such costs may be in part due to restraints or the one-to-one support required to manage the problem behaviour (Tureck et al., 2013). ASD has a sizeable impact on society and treatment of problem behaviour and increased independence that can be achieved with ABA-based interventions may reduce the burden on society (Baxter et al., 2015).

ABA as Treatment Model

Behavioural approaches, such as ABA, that adhere to rigid scientific methodology have been demonstrated as the most common method of treatment for individuals with DD (Weitlauf et al., 2014) because it is the “the only method of instruction which has shown consistent empirically supported improvements in the core deficits of the disorder” (Grey, Honan, McClean, & Daly, 2005, p. 210; see Eikeseth, Smith, Jahr, & Eldevik, 2002; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005) as well as reliable reductions of problem behaviour (Heyvaert, Sanen, Campell, Maes, & Onghena, 2014; Lee, Sugai, & Horner, 1999). Behaviour consultants who use ABA differ in their approach to treating behaviour when compared to others in human services. Rather than relying on the diagnosis or categories of problem behaviour displayed by the individual, behaviour consultants observe the specific behaviour targeted for intervention, create an operational definition, and conduct a functional behaviour assessment (FBA) to determine what conditions maintain the problem behaviour (Matson & Nebel-Schwalm, 2007). The FBA results are then used to select specific ABA interventions to treat the

problem behaviour. These operant techniques can include antecedent or consequence manipulations (Cooper, Heron, & Heward, 2007). Some examples of ABA interventions include differential reinforcement (Allison et al., 2012), extinction (Vollmer & Athens, 2011), functional communication training (Durand and Carr, 1991), positive reinforcement (Bowers, Jensen, Cook, McEachern, & Snyder, 2008), or punishment (Martinez & Betz, 2013).

What is described above is the beginning stage of a FBA, which is used to inform the treatment plan (Koegel, Koegel, & Dunlap, 1996; Griffiths & Watson, 2004). The procedure for completing an FBA is to select a specific target behaviour, determine the functional relationship between the target behaviour and the environment, design an intervention based on behaviour analytic principles based on the assessment, and then monitor progress and make data-based decisions. Broadly, ABA services for problem behaviour can be broken down into three stages: (a) assessment, (b) intervention, and (c) observation and monitoring.

Assessment. ABA treatments are based on an individualized assessment of the problem behaviour and are tailored specifically to the outcome of the assessment, environmental variables, and skills of the caregivers in implementing interventions (Steege et al., 2007). The first step in behavioural assessment is to create an operational definition of the target behaviour that is objective, clear, and complete (Kahng, Ingvarsson, Quigg, Seckinger, & Teichman, 2011). The next step is to determine the function of the problem behaviour using an FBA. Behaviours serve functions that are either socially mediated, in that, reinforcement is delivered by another individual (Cipani & Schock, 2011), or non-socially mediated, where reinforcement is delivered automatically or internally (Iwata & Dozier, 2008).

Interventions based on a FBA have been demonstrated to be more effective than any specific category of intervention (Carr, Robinson, Taylor, & Carlson, 1990). The purpose of an

FBA is to determine the function(s) of the problem behaviour, and use that information to develop effective function-based interventions. There are several different methodologies that can contribute to an FBA. Descriptive assessments can include direct and indirect methods. Indirect methods such as interviews (Kelley, LaRue, Roane, Gadaire, 2011) and questionnaires (e.g., Durand & Crimmins, 1988) are used to identify antecedent variables of the target behaviour. Direct methods (e.g., observations) are used to determine the natural contingencies controlling the behaviour and suggest potential functions of behaviour (Thompson & Borrero, 2011). Together, these descriptive methods are used to identify antecedent and maintaining variables and the conditions in which they occur to formulate a functional hypothesis. Functional analysis, also called an experimental functional analysis (EFA) or a functional analogue (FA), are direct assessments of the function of problem behaviour that include systematic manipulations of antecedent and consequent variables to determine the function of behaviour (Iwata et al., 1994). Functional analyses can be used to confirm functional hypotheses derived from descriptive assessment methods (Hanley, Iwata, & McCord, 2003). These analyses can be completed in the natural environment (in situ) or in a controlled setting depending on the nature of the problem behaviour as well as the need to isolate the assessment from potentially confounding variables. Once the function of problem behaviour is determined, an intervention plan can be developed in order to reduce problem behaviour for individuals with DD.

Intervention. Each intervention is individualized based on target behaviours identified for the individual and their areas of support (Burns, 2015). Further the specific function(s) identified during the FBA must be considered when treatments are selected.

A cookie cutter treatment approach (i.e., if aggression, then time out) may not be effective because the function of behaviour would not be considered (Townsend & Westerfield, 2010).

Interventions can be broadly described as having two objectives: to increase desirable target behaviours and to decrease undesirable target behaviours.

Treatment objectives can be achieved through antecedent modifications, skill development programs, and/or planned consequences that immediately follow the behaviour. Antecedent modifications alter the environment in order to eliminate or reduce the presence of situations associated with high rates of problem behaviour (Cooper et al., 2007). Some examples of antecedent modifications include the removal an event that has previously been associated with the problem behaviour (e.g., avoiding large crowds that typically precipitate self-injury) or interspersing difficult tasks with easy ones (e.g., scheduling a pizza party after a dentist appointment).

Skill development strategies might include teaching functionally equivalent replacement as an incompatible response to problem behaviour. For example teaching functional communication training as a replacement behaviour for self-injurious behaviour (see Durand & Carr, 1991). Skill acquisition programs can also teach skill deficits in academic (e.g., sight word teaching, Kupzyl, Daly, & Anderson, 2011), adaptive (e.g., snack preparation, Bozkurt & Gursel, 2005), and or social skill areas (e.g., engagement in recreational activities, Hughes et al., 2004).

Function based consequences are developed directly from the results of the FBA. A consequence is the stimulus that immediately follows the behaviour. A consequence either: (a) increases or (b) decreases the likelihood of that particular behaviour occurring in the future. A reinforcer strengthens the behaviour and a punisher weakens the behaviour (Miltenberger, 2008). Often, reinforcement and punishment are used in combination. For example, a study by Dickman, Bright, Montgomery and Miguel (2012) used a token economy to increase appropriate vocalizations (i.e., positive reinforcement) and task demands to decrease vocal stereotypy (i.e.,

positive punishment). Problem behaviour may be maintained by negative reinforcement. For example, a child screams to avoid completing homework. The parent decides that homework is unnecessary, and removes the task demand. The removal of the homework is negative reinforcement because the stimulus is taken away and increases the likelihood of screaming in the future to avoid homework.

Observation and monitoring. To determine the effectiveness of behavioural interventions, the target behaviours are observed and measured on a specific dimension (e.g., frequency, duration), which can be used to demonstrate meaningful change (Burns, 2015; Hurwitz & Minshawi, 2012). Behaviour consultants make treatment decisions based on the data collected (Bicard, Bicard, and the IRIS Center, 2012). In order to determine if the intervention is effective, the behaviour consultant takes data on the behaviours and compares any changes to the baseline measurement. If the behaviour changes after the intervention, then a functional relationship can be established between the behaviour and the intervention (Kennedy, 2005). However, in order for this change to be observed, the correct measurement system must be selected.

There are many different types of measurement typically used in research and practice. Continuous measurement such as event recording, duration, and latency, record each instance of the behaviour (Cooper et al., 2007). Event recording records how often behaviour occurs during a specific period of time and is recommended for behaviours that are “discrete” or have a clear beginning and end (Hurwitz & Minshwai, 2012, p. 97). Some examples of behaviours that are recorded by event recording include mands or pinching. Duration is also typically utilized for discrete problem behaviours such as a vocal stereotypy (e.g., Athens et al., 2008). Duration measures the length of the behaviour from its onset to its end (Cooper et al., 2007).

Discontinuous measurement, such as time sampling, may yield inaccurate measurements due to overestimating or underestimating the target behaviour (Hurwitz & Minshawi, 2012).

As described above, ABA is the application of behavioural principles to assess the function(s) of target behaviour and treat it accordingly. The diversity of assessment, treatment, and measurement procedures make implementation of large-scale program evaluation of ABA based service challenging. In addition to the potential range of procedures used in providing ABA-based services, there are different models of service delivery, which have to be taken into account. Some of these services will now be described.

ABA Service Delivery

In Ontario, there are several supports available through the Ministry of Children and Youth Services (MCYS) for children and youth with DD and their families who require ABA-based intervention services (MCYS, 2015). In addition to Intensive Behavioural Intervention (IBI) available for children with ASD through the Autism Intervention Programs, ABA-focused services for children with ASD in Ontario include ABA-based consultation services, and the school support program. Additionally, respite, residential, and employment supports for individuals with DD are available through the Ministry of Community and Social Services (MCSS), who provide funding for the community behaviour management teams in each region (MCSS, 2015).

ABA services can be delivered through direct intervention or consultation mode. In the direct intervention mode (such as the IBI program, and some aspects of the ABA-based consultation services), ABA-trained clinicians meet with individuals and their families to discuss goals and outcomes and then provide intervention directly to individuals or groups. Well-trained and supervised staff typically deliver the direct intervention model. In the consultation model

(such as other parts of MCYS ABA-based consultation services, School Support Program, and MCSS Community Behaviour Management Teams), ABA-trained behaviour consultants systematically train family members and professionals including direct care staff and teachers as “mediators” to implement interventions.

Mediator model of service delivery. The most common type of ABA service in Ontario is delivered in the mediator model or consultation model relying heavily on training natural caregivers/mediators. Mediator training is a process by which a trained behaviour consultant trains a mediator to implement the behavioural intervention. In a best-practice model, the behaviour consultant might use behaviour skills training (BST) model involving seven steps of training the individual to mastery (Parsons, Rollyson, & Reid, 2012). This BST model uses ABA methodology to conduct the training through verbal explanation, written explanation, modeling, role-play, and immediate feedback. The aim of the BST model is to blend skill performance and competency in order to increase the overall quality of work (Reid, Parsons, & Green, 2012). The mediator model is a cost-effective method to train parents in their natural environment (Kaiser, Hancock, & Nietfeld, 2000; Williams, 2000). Mediator training is encouraged by the MYCS as involvement in their child’s programming assists in the generalization and maintenance of the skills that the child is learning and can even reduce parent stress as they learn to cope with challenging behaviours (National Research Council, NRC, 2001). However, there are some challenges within the model that impact program evaluation.

Challenges within the mediator model. Kratochwill and Van Someren (1995) sought to address some common problems observed within the mediator model. The consultation process between a behaviour consultant and mediator includes problem identification, problem assessment, and problem intervention. To be most effective, the behaviour consultant is required

to train the mediator (e.g., the caregiver) to implement the intervention. Despite the availability of the BST model to impart skills on natural caregivers, often, behaviour consultants do not carry out the program alongside mediators. This is problematic because “many staff in human service agencies often do not acquire the skills that the procedures are intended to train” (as cited in Parsons et al., 2012, p. 2).

Treatment integrity and the effectiveness of treatment are positively correlated (Fryling, Wallace, & Yassine, 2012 et al., 2012); therefore outcomes may not be met unless treatment is administered correctly. Thus, when mediators do not adhere to treatment procedures, the effectiveness of treatment can be negatively impacted. There is variability in the quality and type of mediator training that caregivers receive, as many behaviour consultants rely on didactic training (a verbal explanation of the intervention) or leaving a written plan (Kratochwill & Van Someren, 1995), instead of relying on performance-based and competency-based training (Parsons et al., 2012). These differences in training procedures could unintentionally confound the outcomes of mediator-implemented interventions. Mediator training using the BST model can solve both of these issues because it fundamentally addresses areas of treatment adherence (Reid, O’Kane, & Macurik, 2011). In other words, the individual is trained to mastery under the supervision of a behaviour consultant and *in-vivo* practice with feedback takes place (i.e., performance-based and competency-based training as demonstrated in Parsons et al., 2012). Within the BST model, the mediator continues training until treatment adherence reaches a pre-established criterion (Napolitano, Madaus Knapp, Speares, McAdam, & Brown, 2012). The BST model offers opportunities to addresses many of the questions that arise during practical application, making implementation less daunting for caregivers, and increasing the likelihood of adherence. It is important to note that the BST model offers a good training methodology

however the solution to low treatment adherence even after training is to have ongoing observation and monitoring through behaviour consultant supervision (Reid et al., 2012; Parsons et al., 2012).

Because of the lack of standardization, the consulting process itself can influence treatment adherence (Kratochwill & Van Someren, 1995). For example, mediator training that does follow the BST model for training may encounter specific challenges while attempting to implement individualized interventions, leading to lower treatment adherence rates. It is estimated that between 16% and 24% of parent mediators do not carry out the behavioural and medical treatments prescribed for them by medical personnel (Moore & Symons, 2011), therefore altering the effectiveness of the treatment. Research by St. Peter Pipkin, Vollmer, and Sloman (2010) found that mediators do not need to complete interventions with 100% accuracy in order for the intervention to be effective; even 80% treatment integrity was still effective. However, different types of adherence errors can have different impacts on treatment outcomes in ABA. Errors of commission, or when the treatment implementer reinforces the problem behaviour, are more detrimental to treatment effectiveness than errors of omission, where the implementer does not reinforce target behaviour (Fryling et al., 2012; St. Peter Pipkin et al., 2010). St. Peter Pipkin et al. (2010) found that mediators are more likely to commit errors of commission, thereby intermittently reinforcing the problem behaviour. Erroneous or inadvertent reinforcement, especially on a variable schedule, can make behaviour more difficult to extinguish (Ferster & Skinner, 1957).

Another issue that influences adherence in the mediator model is program acceptability (Kratochwill & Van Someren, 1995). A program that is valued increases the acceptability of the mediators and consequently influences their capacity to deliver service. When parents believe

that a program can effect change, are confident in their own skills as a mediator, and have accepted their child in the family and community life, they are more likely to adhere to the program (Moore & Symons, 2011). Parental perception is an important variable to note because unlike static characteristics (e.g., socio-economic status), the dynamic characteristics (e.g., perception of the individual as an effective change agent) can be improved through behavioural techniques (Kazdin, Siegel, & Bass, 1992). Therefore, mediator buy-in is an important aspect to the success of an intervention (Reid et al., 2012).

Measuring Problem Behaviour

While prevalence data typically relates to the presence or absence of problem behaviour in a particular population, prevalence rates do not consider the severity of the target behaviour (e.g., Eisenhower et al., 2005). In practice, the severity of problem behaviour can be quite variable ranging from infrequent life-threatening behaviours (e.g., darting into oncoming traffic) to a constant and pervasive rate of occurrence of an interfering behaviour (e.g., vocal stereotypy), such that it impacts everyday life and learning (Durand, 1990). Severe problem behaviour such as self-injurious behaviour or aggression may cause permanent damage and hospitalization (Durand, 1990) and may explain the high rates of psychotropic medication among individuals with ASD (Aman, Lam, & Collier-Respin, 2003). On the contrary, other problem behaviour may significantly disrupt life quality but it may not be severe, such as hand flapping (Didden et al., 2012). Both ends of the continuum can impact life quality. Less severe but equally pervasive problem behaviours include resistance to transition, non-compliance, tantrums, and repetitive behaviours such as toe walking or rocking (Durand, 1990).

The topography of behaviour, on the other hand, refers to “the physical form or shape of a behaviour” (Cooper et al, 2007, p. 707). Common topography measurements such as the

Problem Behavior Questionnaire (PBQ; Lewis, Scott, & Sugai, 1994) indicate the type of problem behaviour but do not specify to what intensity it occurs. The severity of problem behaviour is often assessed across dimensions such as frequency, intensity, and duration (Wilczynski, 2011).

In practice, data collection systems are designed to incorporate the dimensions of behaviour most relevant to the goal of intervention. However, in research and evaluation, a common measure of problem behaviour is required. Typical measures of problem behaviour focus on one or more dimensions for multiple topographies, but do not offer in-depth evaluation of a single behavioural target (e.g. Behaviour Problems Inventory BPI; Rojahn et al., 2001).

Program Evaluation

Program evaluation is a process used to determine if a service or intervention (“program”) that is being applied leads to the desired outcomes (Posavac & Carey, 2007). Program evaluation can be used to determine how a specific program is operating. Program evaluation enables providers to determine the program effectiveness, and whether it is (a) achieving desired outcome, (b) needs to be modified to achieve desired outcome or, (c) is not able to/ designed to achieve desired outcomes and needs to be stopped. In addition to examining effectiveness, program evaluation can also be used to determine the efficiency and cost effectiveness of a particular program. In this way, “program evaluators gather information to help people improve their effectiveness, to assist administrators to make program-level decisions, and to enable interested parties to examine program effectiveness” (Posavac & Carey, 2007, p. 10).

Posavac and Carey (2007) described the importance of program evaluation and reviewed several benefits of its use. Program evaluation can address gaps within the service model by

determining the degree to which needs are met, and to identify unmet needs (Gaber, 2000).

Program evaluation can verify that the planned programs have been put into practice. Program evaluation helps to define and refine the desired program outcomes into measureable terms and permits examination of the program outcomes. If more than one approach is used to obtain the desired outcomes, program evaluation can be used to compare the outcomes of each approach and determine which one is the most effective. Where needs remain unmet, program evaluation can be used to determine the need for additional services or the re-allocation of services.

Program evaluation can detect unplanned side effects that may impact the outcome. Program evaluation can also test assumptions made about the program, such as, the assumption that the program directly leads to the outcome, and that other factors are not responsible for the observed change. Program evaluation can be used to test this assumption by carefully documenting program components and determining their contribution to the desired outcome. Program evaluation can also provide information necessary to maintain and improve the quality of service to produce meaningful outcomes (Schalock & Verdugo, 2012). In this age of accountability, program funding is becoming dependent on consistent demonstration of expected outcomes.

Quality assurance ensures that this process takes place. Posavac and Carey (2007) summarize:

Program evaluation can contribute to the well-being of society only if evaluators successfully meet their obligation to help government agencies and private organizations focus on important needs, plan effectively, monitor carefully, assess quality accurately and justly, nurture improved practices, and detect unwanted side effects. (p. 7)

The feedback derived from program evaluation is: (a) formative, because it helps to improve existing services or outcomes and (b) summative, as it can help to determine if an program should start, continue, or stop based on identified outcomes (Posavac & Carey, 2007).

This feedback assists in quality assurance, because it is used to monitor ongoing quality throughout the services or programs being delivered (Posavac & Carey, 2007).

The Need for Systematic Measurement of Problem Behaviour

Program evaluation of ABA treatment as a whole is not common within behaviour analysis (Weitlauf et al., 2014). Some programs use clinical measures of behavioural symptoms to address this need (e.g., My child physically attacks people; Achenbach, 1978); however, these measures may not be sensitive to the discrete behavior changes that are targeted by ABA-based services. Successful ABA programs result in relatively rapid behavioural changes; therefore, effective program evaluation must be frequent and ongoing in order to capture these outcomes (Posavac & Carey, 2007). One of the barriers to effective program evaluation for ABA-based services is the degree to which treatment targets, assessments and interventions are individualized making comparisons of interventions and outcomes across clients difficult at best. The variability inherent in individualized programming makes program evaluation for the agencies delivering these services quite challenging as the targets of intervention, types of measurement, and specific procedures differ across cases (Posavac & Carey, 2007; Weitlauf et al., 2014, p. ES-20).

Program evaluation of changes in behaviour is difficult to complete (Posavac & Carey, 2007). A second challenge of program evaluation is the diversity of approaches for measuring problem behaviour. Existing measures (e.g., BPI, Achenbach System of Empirically Based Assessment-Child Behavior Checklist, etc.) may be inadequate for specific ABA intervention planning because they survey all topographies of problem behaviour, rather than provide multiple ratings of the dimensions of a particular behaviour or set of behaviours (Eisenhower et al., 2005). Measurement may not be sensitive to minor changes that have occurred during

programming (Burns, 2015). McClintock et al. (2003) stated that within the field of developmental disabilities, there is a “lack of data on incidence, prevalence, and chronicity of challenging behaviour” (p. 3). In addition, there are few available measures that incorporate multiple dimensions of severity across problem behaviours that occur across and within settings. A measure needs to capture changes that have occurred; otherwise it can produce inaccurate information (Wilczynski, 2011). Therefore a measure must be sensitive to change across various dimensions that indicate the severity of problem behaviour. In addition, the desired outcomes must be defined and measureable otherwise measurement of outcomes is not possible. Posavac and Carey (2007) outlined the logic of the need for measurement in this way: “Without measuring need, planning cannot be rational; without effective implementation, good outcomes cannot be expected; and without achieve good outcomes, there is no reason to worry about efficiency” (Posavac & Carey, 2007, p. 10). Program evaluation cannot be completed unless there is a consistent measure that can assess outcomes within and across agencies. Given the cost of service delivery, there is a substantial need to validate measures designed for program evaluation to determine whether expected outcomes are being met through treatment (Posavac & Carey, 2007).

Existing Measures of Problem Behaviour

Most outcome studies for behavioural interventions (e.g., parent training studies) have evaluated the severity of problem behaviour through time-series behavioural data measures. During the assessment phase of ABA treatment, a parent, caregiver, or teacher who sees the individual in a variety of settings, may be asked to fill out a measure to indicate problem behaviour. These measures can be used to indicate problem behaviour at baseline and during the intervention. Information gained from each measure should be used in conjunction with other

data in order to inform the child's treatment, a process that takes clinical expertise (Rescorla, 2005). Some disadvantages of the common measures used to indicate the severity of problem behaviour will be described.

The ASEBA-CBCL (Achenbach, 2009) is the new version of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000). This 99-item form measures the respondent's perception of child behavioural and emotional problems in young children aged 1.5 to 5 years that have occurred within the last 2 months. Parents, teachers, or caregivers use a 3-point Likert scale to indicate whether the item is 0, not true, 1, somewhat or sometimes true, or 2 very true or often true. This measure indicates sleep problems but more relevant to the current study is that it yields a 'total problem score', which can be broken into 2 internalizing and externalizing problem scales. The 6 subscales of these categories include: (a) emotionally reactive, (b) depressed/anxious, (c) somatic complaints, (d) withdrawn, (e) attention problems, and (f) aggressive behaviour. Totals within each syndrome indicate whether the problems are within the normal, borderline, or clinical ranges. The CBCL displays good psychometric properties (Achenbach & Rescorla, 2013) including reliability and validity (Rescorla, 2005). Behavior problems represented by the CBCL have been highly correlated with other measures such as the Conners (Landy & Bradley, 2013; Rescorla, 2005). CBCL and its form variations have been used in a variety of studies that assess problem behaviour (i.e., Baker et al., 2003; Eisenhower et al., 2005; Feldman et al., 2000). Eisenhower et al. (2005) mentioned one limitation of the CBCL is that it is not syndrome specific and does not account for specific behaviours of ASD such as "self-injury, insistence on routine, social avoidance, dysregulated sleep and waking cycles" (p. 11). The ASEBA-CBCL may not provide an accurate description of problem behaviour based on

syndrome specific traits. Despite its limitations, CBCL is a cost-effective and user-friendly measure (Rescorla, 2005).

The Developmental Behaviour Checklist (DBC; Einfeld & Tonge, 2002) is a 96-item measure used to assess behaviour and emotional problems of children with DD. Parents or teachers respond to questions regarding problem behaviours that have occurred over the preceding 6 months using a 3-point scale to indicate whether they knows the item to be 0, as not true, 1, as somewhat or sometimes true or 2, as very true or often true. The DBC yields a 'total behaviour problem score' with 5 subscales which include: 1) disruptive/antisocial, 2) self-absorbed, 3) communication disturbance, 4) anxiety, and 5) social relating. The DBC has poor interrater reliability between parents and teachers but high concurrent and construct validity as well as high internal consistency (O'Brien, Pearson, Berney, & Barnard, 2001). It has high correlations with alternative measures that assess problem behaviour exhibited by children with IDs (Einfeld & Tonge, 2002). The DBC has been used as a measure to represent associations between problem behaviour and parent life quality (e.g., Herring et al., 2006) as well as to assess problem behaviour as it relates to severity of ID (e.g., Dekker, Nunn, Einfeld, Tonge, & Koot, 2002; Koskentausta, & Almqvist, 2004). The DBC is sensitive to change over time (Harris, 2006; Tonge & Brereton, 2011); however, neither the severity nor specific topographies of problem behaviour are represented.

The Behavior Problem Inventory-01 (BPI-01; Rojahn et al., 2001) is a 49-item measure that provides an indication of the frequency and severity scale of problem behaviours exhibited by individuals with ID aged 14-91 years that have occurred within the last 2 months. Parents or caregivers rate the severity (1 = slight, 2 = moderate or 3 = severe) and frequency (0 = never, 1 = monthly, 2 = weekly, or 4 = hourly). The BPI-01 produces a total as well as sub-scores for: (a)

self-injurious behavior, (b) stereotyped behavior and, (c) aggressive/destructive behavior. Rojahn et al. (2012) confirmed that the BPI-01 has strong convergent and discriminant validity.

Reliability was found to be strong within this measure (Rojahn et al., 2001). The BPI-01 has been compared to other measures of problem behaviour (e.g., the ABC in Rojahn et al., 2013 and Autism Spectrum Disorders-Behavior Problems for Intellectually Disabled Adults in Rojahn, Wilkins, Matson, & Boisjoli, 2009) in order to assess the validity against other measures and has found similar results. Still, Rojahn and colleagues (2013) advised that the BPI-01 should be used in conjunction with other problem behaviour measures to provide a detailed description of the problem behaviour. One advantage is that two dimensions of problem behaviour are represented by the BPI-01 whereas other measures have only indicated the presence or absence of problem behaviour (e.g., CBCL or DBC).

Gaps in Existing Measures of Problem Behaviour

Currently, there is a paucity of measures that can be used to assess the severity of a range of specific behaviours or any changes that occur across time. The BPI-01 measures two dimensions of problem behaviour (i.e., frequency and intensity). However, there are other dimensions that can reflect problem behaviour (i.e., duration and discrimination). It is important for problem behaviour to be represented as accurately as possible and if the frequency of it is not a problem but the discrimination of it is (i.e., where it occurs - *public* masturbation), then only a single incident can have problematic outcomes. Hurwitz & Minshawi (2012) recommend developing a Likert scale that depicts the intensity and severity of the problem behaviour to assist in providing an accurate measurement of its scope. Furthermore, it is important for the scale to be objective with clear anchors such as a 1 indicates no blood drawn to a 10 indicating the presence of bone damage (e.g., the Self-Injury Trauma Scale Iwata, Pace, Kissel, Nau, &

Farber, 1990). These anchors can provide some guidance to increase reliability across raters, but they are not sufficient to guarantee reliability as other issues (e.g., different environments, different opportunities for observation) may bias responses. Adjusting the wording and phrasing of the questions on the measure can increase the reliability of a measure, and if it remains low, further analyses can help to determine if additive or within-measure systematic errors have been made (Viswanathan, 2005).

The ASEBA and the DBC measure the presence or absence of many different problem behaviours, but do not provide opportunities for further ratings of severity. This is problematic because these measures do not account for changes in the severity of the problem behaviour. Wilczynski (2011) warns that measures that are not sensitive to the target behaviour or the change over time can produce “inaccurate information” (p. 56) for practitioners. Many problem behaviour measures do not demonstrate a sensitivity to change over time. This means that specific changes in problem behaviour may not be reflected through the measure. The DBC is sensitive over time (Harris, 2006; Tonge & Brereton, 2011), but it does not measure any specific dimensions of problem behaviour.

The Need for Measures that are Socially Valid

In addition to the development of a reliable measure of problem behaviour that is sensitive to change, the measure must also be socially validated. A measure may be efficacious, effective and efficient, but if it is not well liked, it will not be implemented (Wolf, 1978). Social validity is “the extent to which target behaviours are appropriate, intervention procedures are acceptable, and important and significant changes in target and collateral behaviours are produced” (Cooper et al., 2007, p. 704). Montrose Wolf (1978) introduced the term and concept of social validity to the field of ABA and proposed three ways that it can be

assessed: (a) the significance of behavioural goals, (b) the appropriateness of behavioural procedures, and (c) the social importance of the results. First, the goals and procedures chosen must be relevant to the desired lifestyle change in order to be socially valid (Wolf, 1978). Consequently, if a treatment is not considered to be beneficial to the clinical community, it will not be a popular choice among consumers. This will in turn, lead to social invalidity (Schwartz & Baer, 1991), which is not only a disagreement of program value but also compromises the “implementation and dissemination” of such program (Gambrill, 2012, p. 123). Second, the behavioural procedures must be appropriate to be socially valid. This includes the degree to which an intervention is acceptable or liked (Baer, Wolf, Risley, 1987). Social validity is especially important to consider within the mediator model because if the mediators agree with the goals and procedures of the intervention, treatment acceptance has a direct effect on intervention outcomes (Fryling et al., 2012). The effort, time, possible discomfort, and ethics are all factors that influence how change is produced (Wolf, 1978). Determining the significance of behavioural goals within family expectations is therefore a crucial step in the development of an individualized plan (Turnbull & Reuf, 1997). Third, the results of the intervention must be socially important. Measures to reduce the severity of problem behaviour must be socially valid in two ways: (a) the target behaviour change must be socially significant for the individual in their environment, and (b) the changes that are made must be socially meaningful. For example, if vocal stereotypy is reduced from 75% to 10% in the school setting this reduction might be socially significant. However, if 10% is still too disruptive for the classroom setting, and results in time spent outside of the classroom, the reduction is not socially meaningful.

Social validity should address customer satisfaction regarding the outcome of the predicted behaviour change and any possible unwanted side effects in order to influence the

future implementation of such intervention (Schwartz & Baer, 1991). There are some guidelines on customer satisfaction data to promote social validity. Wolf (1978) stated that social validity data should support objective data. That is, in addition to being effective, the outcome must also be acceptable to recipients. Thus the aim is to incorporate treatment that is both effective and acceptable (Wolf, 1978). Schwartz and Baer (1991) recommended that in order to accurately capture the differential responses of social validity, a 7-point Likert scale should be used in conjunction with an area to provide open feedback.

Social validity is an integral part of research (Hurwitz & Minshawi, 2012; Schwartz & Baer, 1991) and clinical application because it influences the selection of behaviour targeted for reduction and whether the result of intervention is meaningful. Reducing problem behaviour is an important social value. Not only is important for the severity of problem behaviour to reduce, this change must be meaningful in order to lead to an improved life quality (Feldman et al., 2002). An outcome that displays statistical significance does not necessarily equate meaningful social change (Wilczynski, 2011). The change that is produced must be socially meaningful otherwise “society will be less likely to use our technology, no matter how potentially effective and efficient it might be” (Wolf, 1978, p. 206). He summarized the ultimate purpose underlying the practice of ABA is to produce meaningful change that improves the life quality of individuals. Best practice of ABA includes reducing problem behaviours in order to improve life quality (Horner et al., 2002).

In previous research, Feldman et al. (2002) examined social validity as a component on the field effectiveness of positive behaviour support for individuals with DD who displayed severe problem behaviours. They used locally developed social validity questionnaires assessing the parent’s perception of the severity of the problem behaviour, which became known as the

FIDD in the clinical services in which it has been adopted. The FIDD measured severity across four dimensions: (a) frequency, (b) intensity, (c) duration and (d) discrimination. In their study, pre- and post- intervention data demonstrated a substantial decrease across all dimensions of parent perceived severity of problem behaviour. In addition, aspects of quality of life were separately rated and showed gains in most areas including: (a) participant's engagement in daily activities and routines, (b) opportunities for community integration, (c) peer relationships, (d) learning, (e) family stress, (f) family attending social events outside the family home and (g) family hosting social functions at home. Finally, they measured consumer satisfaction with the intervention package, (a) acceptability, (b) effectiveness, (c) intrusiveness, (d) quality, (e) satisfaction, and (f) recommendation to others, range from a 4.2 to a 4.9 within a score range of 1 to 5. This study is valuable to the field because it assessed a variety of gaps within ABA regarding various dimensions of problem behaviour in addition that were considered socially valid by parent participants. This 4-item measure (also used in Feldman & Werner, 2002 to measure child behaviour problems) was the precursor measure to the development of the Target Behaviour Severity Scale (TBSS; Condillac, 2009d) that is used in the present study. Since the Feldman et al (2002) publication, few studies have included measures of social validity relating to the reduction of problem behaviours.

Rationale

Purpose

The TBSS is designed to provide ratings of the dimensions of a specific problem behaviour in order to examine social validity of behaviour change. This information can be used in conjunction with other measures to evaluate the efficiency and effectiveness of behaviourally-based services, by describing outcomes for particular groups, interventions, or assessment

procedures. The purpose of this study is to examine the reliability and validity of the TBSS in a sample of behaviour consultants and caregivers of individuals with DD.

Research Questions

1. Does the TBSS have acceptable internal consistency? The first hypothesis is that the TBSS will have acceptable internal consistency when completed by behaviour consultants and caregivers.
2. What is the inter-rater reliability between behaviour consultants and caregivers on the TBSS? The second hypothesis is that behaviour consultant's and caregiver's responses will be positively correlated; however there will be mean differences across raters.
3. What are behaviour consultant's and caregiver's opinions of the face validity and usability of the TBSS? The third hypothesis is that the TBSS will be described as a useful and easily understood measure of the severity of problem behaviour.

Method

Recruitment

Agencies that were a part of earlier phases of the development of the measure were invited to participate in the current study. These participating agencies deliver government funded behavioural intervention services to children and adults with DD in Ontario. Information sessions were held with managers and behaviour consultants to provide an overview of the Measures Study including the background, rationale, procedure, consent process, and review of the measures. Note that data for this study were collected in the context of a larger study focused on the development and field-testing of program evaluation measures for ABA-related clinical programs being conducted by Dr. Rosemary Condillac. For this aspect of the study, consent forms and client packages were left with each behaviour consultant, with instructions to review

the consent form and if they decided to participate, to follow the instructions to recruit clients. Of the 24 behaviour consultants who attended the information sessions, 11 participated in the study. Each behaviour consultant who agreed to participate was asked to approach two to five clients with a range of problem behaviour for the study. Behaviour consultants were asked to use a phone call script to recruit caregivers of individuals with DD. If consent was obtained, the behaviour consultant completed the TBSS regarding the behavioural issue being treated. During the next scheduled visit, the behaviour consultant read the in-person consent script and initialled as witness that consent was obtained, but did not record the names of the caregivers or individual with DD. The 11 behaviour consultants approached a total of 40 caregivers to participate. Eleven caregivers did not consent to participate when asked over the phone. There were 29 participants who consented to participate via phone script, with two declining when asked at the visit. A final sample of 27 caregivers agreed to participate when consent was confirmed at the visit. No participants or behaviour consultants withdrew after data had been collected.

The 11 behaviour consultants and 27 caregivers completed the TBSS within a few days of each other to avoid confounds due to changes in behaviour over time. The TBSS was completed at the visit while the behaviour consultant was present. The behaviour consultant ensured that the operational behaviour was copied onto the TBSS before participating caregivers completed the TBSS for the target behaviour and returned it to the behaviour consultant. The caregivers could ask the behaviour consultant questions if they had any difficulty with the measure. The caregivers also completed a measure relating to the impact of the behaviour on life quality, which is part of the larger project described earlier. The behaviour consultant placed the caregiver's completed measures in a sealed envelope along with their own completed TBSS,

their own consent form, and the consent script with their witness that verbal consent was obtained, from the caregiver, and returned it to the designated person at their agency.

Participants

Consultants. Eleven behaviour consultants from four clinical teams across two agencies participated in this study.

Caregivers. The sample included 27 caregivers supporting 27 individuals receiving publicly funded behavioural services for one or more target behaviours. Participants were involved with a behaviour consultant but could have been in the intervention phase or the assessment phase. This was intentional because the measure was designed for use throughout assessment and treatment phases. Of the 25 caregivers, 16 disclosed their relationship to the person with DD. The caregivers were mostly family members, with 44% parents ($n = 11$) and 4% grandparents ($n = 1$). Caregivers also included supervisors that provided care for the individual in day programs (4%, $n = 1$) or residential settings (12%, $n = 3$). Many participants reported gender-neutral relationships (44%, $n = 11$). Only 15 caregivers reported their gender (8% males and 52% females).

Person with DD. The individuals with DD who displayed the problem behaviour included 12 children and three adults, while the ages of 12 participants were not reported. No other demographics were obtained.

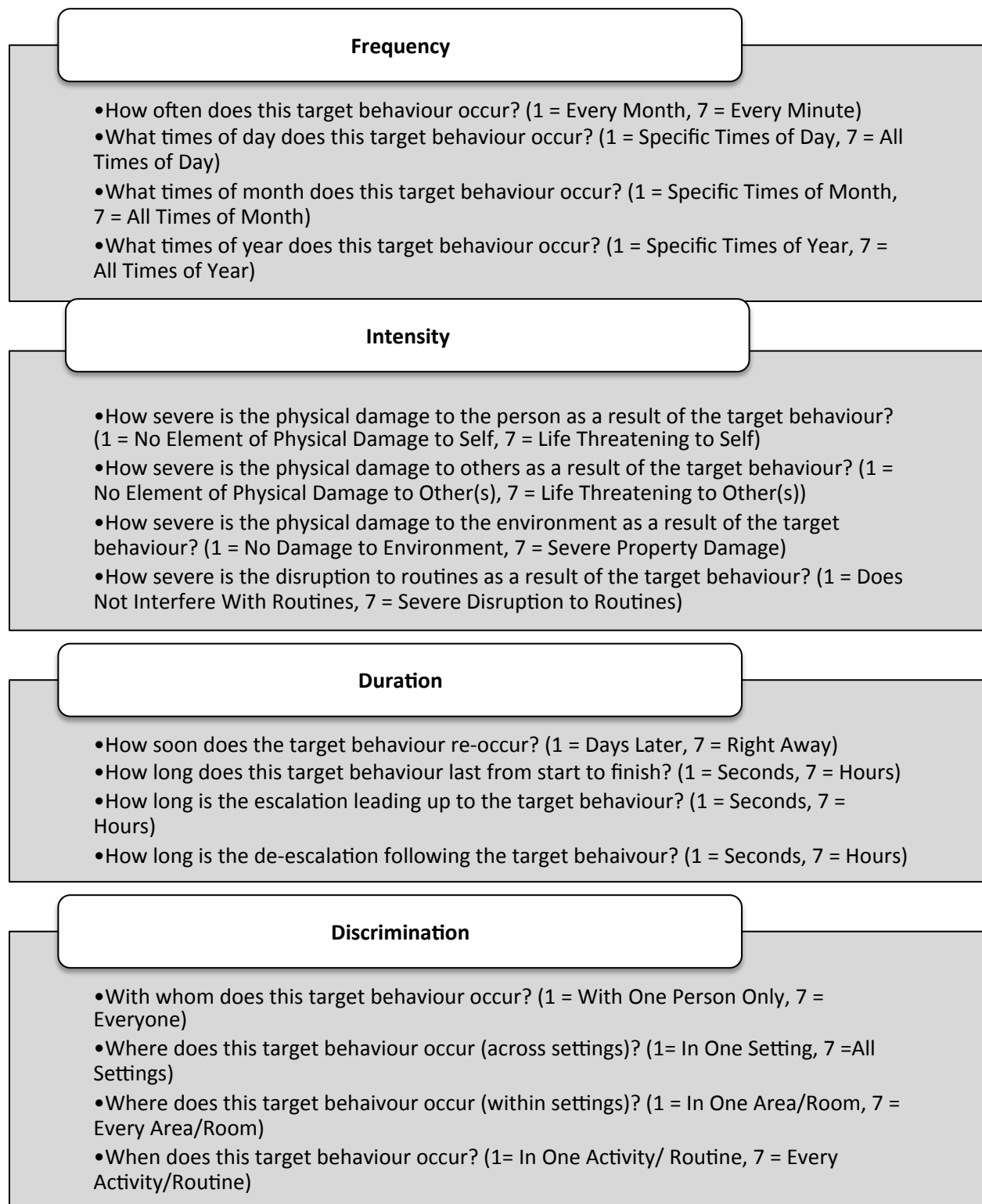
Measures

Figure 1. Breakdown of TBSS questions across dimensions of problem behaviour.

TBSS. The Target Behaviour Severity Scale (TBSS; Condillac, 2009d, see Figure 1) was designed for use by behaviour consultants and caregivers to measure the severity of problem behaviour across four dimensions: frequency, intensity, duration, and discrimination, as an expansion of the 4-item scale used in previous research by Feldman et al. (2002). The combination of the four scales represent a ‘total severity score’. The 16-item questionnaire has four questions per dimension. During the initial stages of development face validity and feedback were established across four teams of behaviour consultants ($n = 39$), and minor revisions were completed that resulted in the current version of the TBSS used in this study (Condillac, 2009d). Items are each scored on 7-point Likert-type scale that is anchored with statements specific to the dimension of behaviour being rated (e.g., One intensity question has the following anchors: 1 = does not interfere with routines, 3 = minor interference/disruption, 5 = major interference/disruption, 7 = severe disruption to routines). The organization and content of the TBSS is represented in Figure 1. Each dimension of the TBSS will now be described.

Frequency. Frequency indicates how often the target behaviour occurs. The amount of time can fluctuate between small increments of time (i.e., per second) or large increments of time (i.e., per month). This dimension was included because the duration of a target behaviour may be brief it may occur so frequently that it can be a problem behaviour. For example, saying, “Hello!” at the beginning of a conversation is socially acceptable. However if “Hello!” is said 10 times within 10 minutes during a conversation, this behaviour turns into problem behaviour. The frequency dimension of the problem behaviour is represented by occurrences each day, month, and year.

Intensity. Another important dimension of problem behaviour is intensity, which refers to how severe it is. Intensity differs from frequency because a behaviour may occur infrequently an

cause severe damage or conversely, occur often but cause minor damage (Hurwitz & Minshaw, 2012). For example, head banging might occur monthly but may lead to brain damage after a single blow. Furthermore, reductions in frequency do not necessarily mean reductions in intensity (LaVigna & Willis, 2012). The TBSS measures intensity through the extent of physical damage to the individual, others, and the environment as well as the disruption of routines as a result of the target behaviour.

Duration. The duration of problem behaviour is the length of time between the onset and end of the target behaviour. The TBSS addresses this component by this question, “How long does the target behaviour last from start to finish?” Duration can also include the time immediately before the problem behaviour occurs, when precursor behaviours might be exhibited. “How long is the escalation leading up to the target behaviour?” Duration can be used to measure the length of a tantrum as well as how soon afterwards a second tantrum occurs. This aspect of duration is captured by the following question: “How soon does the target behaviour re-occur?”

Discrimination. It is important to determine the conditions under which, when, and how the problem behaviour is occurring. Problem behaviour can vary greatly across environment, materials, and individuals. For example, a child may be able to read a book with a teacher but not their parents. Alternatively, an individual may only use the bathroom in the home and not in public settings. These concepts are addressed in questions such as “With whom does this target behaviour occur?” or “Where does this target behaviour occur (across settings)?” Some behaviours need to occur in the presence of different people and settings (generalized), such as communication skills, whereas other behaviours should occur only under particular circumstances, and not others (discriminated). For example, disrobing in public is an example of

a discrimination failure in that the behaviour occurs in situations other than those where it is acceptable as disrobing should occur only in private settings rather than at school, or at the mall.

Consultants and caregivers were asked to respond to each item using a 7-point Likert scale. After completing the 16 items, participants were asked to complete some additional questions in order to provide feedback on the face validity, clarity, and general feedback on the usefulness of the tool. See Figure 2 below.

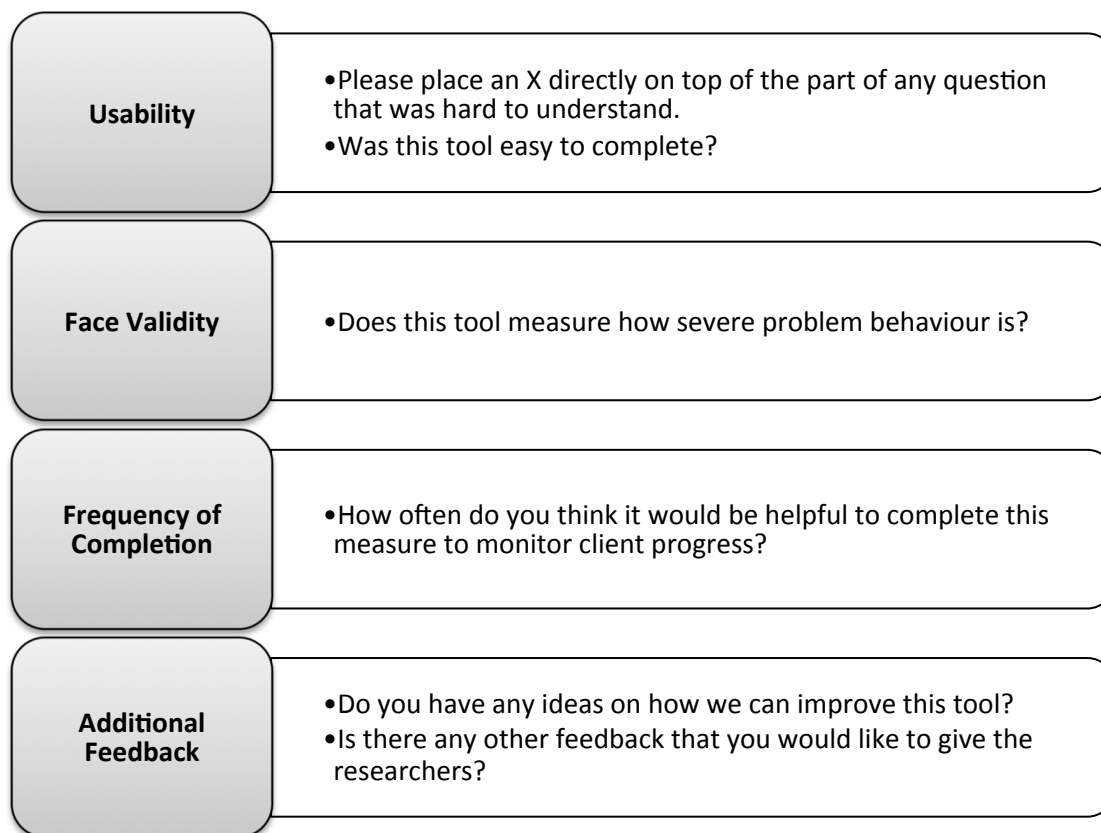


Figure 2. Supplementary Feedback Questions From the TBSS.

Procedure

The following procedures were completed.

1. Behaviour consultants were recruited through participating agencies and asked to recruit caregivers of participants with DD who displayed problem behaviour. Behaviour

consultants were asked to recruit caregivers irrespective of treatment stage (i.e., assessment, treatment, or observation and monitoring).

2. To minimize pressure on the potential participants, behaviour consultants described the study using a phone script, clearly outlining that participation was voluntary and would have no bearing on current or future services. If caregivers provided consent, behaviour consultants initialed the witness section of the script, and completed the TBSS.
3. At the next scheduled visit, the behaviour consultant reviewed the consent script in person with the caregiver. If consent was confirmed, the behaviour consultant provided the caregiver with the TBSS to complete during the visit, ensuring that the same operational definition of the target behaviour used for both raters.
4. Once the caregiver completed the measures, the behaviour consultant sealed the completed TBSS in an envelope with a copy of the witnessed consent script and the behaviour consultant's completed TBSS.
5. This data package was returned to Dr. Condillac's lab for safe storage and analysis. The consent forms for each behaviour consultant were separated from the data to ensure anonymity.
6. When opened, each item within the package was numbered with a study identification code and entered into the database.
7. Analyses were performed using SPSS 22.

Results

Final Sample

As reported above, 11 behaviour consultants and 27 caregivers participated in the current study. Behaviour consultants completed an average of two TBSS measures ($SD = 1.6$, range = 1-

5). An initial examination of missing data on the TBSS measure revealed that two specific caregivers accounted for 80% of the missing data (7/16 and 5/16 missing questions). No discernable or systematic patterns were found for missing data for these two participants: one appears to have skipped an entire page, while another, skipped over questions, but answered some on each page. Replacement of the data for these participants would have potentially increased measurement error. Therefore a decision was made to remove the individuals from the sample for analysis. After these outliers were excluded, there were a total of three missing data points, which accounted for less than 1% of the total scores entered. One participant skipped question 16, while another participant skipped questions 13 and 14. Tabachnik and Fidell (2001) assert that any substitution method chosen will yield similar results; however mean substitution provides the “best guess about the value of the variable” (p. 62). Hence, the three missing values were substituted with group mean scores. After missing data analysis and replacement, 25 caregivers and 11 behaviour consultants were used in the following analyses. In terms of the supplementary feedback section, six out of eleven behaviour consultants provided responses on the face validity, usability, and additional feedback on the TBSS; therefore the results should be interpreted cautiously. Twenty-four caregivers provided some supplementary feedback albeit to different extents. Some caregivers responded to the close-ended questions ($n = 22$) while others did not provide any written feedback ($n = 13$).

Target Behaviours Rated on the TBSS

Participants were asked to describe the target behaviour that they were rating on the TBSS. Behaviour consultants ensured that the caregivers rated the same target behaviour(s) that they had rated. The target behaviours, as described by the participants, were coded into specific behavioural descriptions that were based on Atkinson and Feldman (1994) then grouped into the

broader categories (e.g., physical aggression, property destruction, disruptive behaviour). The frequencies of behaviours in each category are presented in Table 1. Dangerous behaviours were the focus of the TBSS ratings for more than half (56%) of the sample. Tantrums were the focus of the ratings for 20% of the sample and disruptive behaviours and skill deficits were each the focus of ratings for 12% of the sample. Almost all raters (88%) selected target problem behaviours to decrease, while the two remaining raters selected behaviours to increase; the latter scores were reversed to ease interpretation of the results.

Table 1

Target Behaviours Rated on the TBSS

Categories of Target Behaviours	<i>N</i> = 25	%
Dangerous behaviour	14	56
Physical Aggression With or Without Outbursts	5	36
Property Destruction With or Without Outbursts	4	28
Two or more Dangerous Behaviours	3	21
Self-injury	2	14
Tantrums	5	20
Disruptive Behaviour	3	12
Skill Deficits	3	12

Note. Broader categories are boldface.

Consultant's and Caregiver's Responses on the TBSS

The means, standard deviations, and ranges for the items and scales on the TBSS were examined separately for behaviour consultant and caregiver responses. Individual item ratings ranged from 1 to 7, with higher numbers indicating higher severity. Dimension total scores were calculated by summing the responses to the four items corresponding to each dimension. The lowest possible score within a dimension would be 4 (if each item rated as 1) and the highest possible score would be 28 (if all items rated as 7). The total severity score was calculated by summing the responses to the 16 items. Across the dimensions, the lowest possible total severity score would be 16 (if all items rated as 1) and the highest possible score would be 112 (if all

items rated as 7). A summary of the item scores, dimension totals, and total severity scores are presented in Table 2. The summary of the behaviour consultants' responses will be reviewed first, followed by the summary of the caregivers' responses.

Table 2

Summary of Consultant's and Caregiver's Responses to the TBSS

Variable Name (<i>N</i> = 25)	Consultants			Caregivers		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Frequency Total	17.52	5.80	8-26	18.56	4.77	7-24
Occurs	2.88	1.24	1-5	2.84	1.14	1-5
Time of Day	4.16	1.80	1-7	4.48	1.98	1-7
Time of Month	5.20	2.14	1-7	5.64	2.08	1-7
Time of Year	5.28	2.15	1-7	5.60	1.96	1-7
Intensity Total	12.24	6.88	5-28	12.38	6.47	5-28
Damage to Individual	2.88	2.05	1-7	2.92	2.02	1-7
Damage to Others	2.64	2.04	1-7	2.38	1.75	1-7
Damage to Environment	2.72	2.01	1-7	2.96	2.21	1-7
Routine Disruption	4.00	1.58	1-7	4.13	1.67	1-7
Duration Total	14.72	4.15	7-23	15.36	3.87	8-23
Reoccur	2.92	1.73	1-6	3.24	1.72	1-6
Last	4.24	1.23	2-7	4.52	1.19	2-7
Escalation	3.56	1.53	1-7	3.52	1.50	1-6
De-escalation	4.00	1.73	1-7	4.08	1.22	2-6
Discrimination Total	15.52	4.99	7-26	18.28	5.81	8-28
Whom	4.00	1.80	1-7	4.92	1.85	1-7
Across Settings	3.48	1.56	1-7	4.44	1.56	1-7
Within Settings	4.04	1.84	1-7	4.48	2.02	1-7
When	4.00	1.47	1-7	4.44	1.47	2-7
Total Severity Score	60.00	13.12	39-84	64.52	13.06	38-86

Note. Dimension totals and total severity score are boldface.

Consultants' ratings and scores. For each dimension, the behaviour consultants' item level ratings will be described, followed by the total scores. After describing the results for each dimension of severity, the total severity score will be reviewed.

Frequency. A review of the four items that are designed to measure the frequency of problem behaviour revealed that there was a full range of responses from 1-7 provided by the behaviour consultants for three questions and a slightly restricted range (5 out of a possible 7) for

one question. The means for each of these questions showed an upward trend from the general question of how often the behaviour occurs to the questions rating frequencies across days, months and years. The mean total frequency rating score was 17.52 out of a possible 28.

Intensity. There were four items designed to measure the intensity of the target behaviour that included a full range of scores (1-7) reported by behaviour consultants. There was a relatively stable trend for the means for questions on physical damage to the individual, others, and environment. Disruption to routine, however, had a mean rating of 4, and stood apart from the other mean ratings of intensity. The total mean intensity rating score was 12.24 out of a possible 28.

Duration. There was a slightly restricted range for two of the items. One question asked how often the behaviour reoccurs and no behaviour consultants used the most severe indicator (i.e., that the behaviour occurs right away). Similarly, no behaviour consultants reported that target behaviour lasts only seconds from start to finish. Therefore, the ranges of scores for these questions were 1-6 and 2-7, respectively. There was a somewhat variable trend across means where how often the behaviour occurs had a mean of 2.92 whereas how long the behaviour lasts had a mean of 4.24. How long a behaviour lasts and the escalation time to the target behaviour also followed the variable trend with mean scores of 4.24 and 3.56, respectively. The total mean duration rating score was 14.72 out of a possible 28.

Discrimination. Four items designed to measure the discrimination of problem behaviour revealed a full range of response from 1-7 provided by the behaviour consultants for each of the questions. The means for each question showed a stable trend, all near a mean rating of 4 or close to it with the exception of the across settings mean which stood apart from the other scores

with a mean of 3.48. Out of a possible 28, the total mean discrimination severity rating was 15.52.

Consultant's total severity score. Behaviour consultants' mean total severity score was 60 ($SD = 13.12$) with a range from 39 to 84 out of a possible 112. There was a variable trend across dimension totals. The frequency total mean score was distinct from the other dimension totals at 17.52. Similarly, intensity also differed from other dimensions at 12.24 out of a possible score of 28. Duration and discrimination, however, followed the most similar trend.

Caregiver's ratings and scores. Caregiver TBSS ratings will be described first within each dimension follow by the total severity score.

Frequency. There were 4 items designed to measure the frequency of the target behaviour which included a slightly restricted range of scores (5 out of a possible 7) for the how often a behaviour occurs question. There was an upward trend for the means of the frequency questions beginning with how often the behaviour occurs to the questions rating frequencies across days, months and years. The mean frequency total score was 18.56.

Intensity. There were 4 items designed to measure the intensity of the target behaviour that included a full range of responses from 1-7 provided by the caregivers. The means for each of these questions included a slightly variable trend. In particular, routine disruption stood out from the rest with a mean of 4.13. The mean caregiver response for damage to others was another item that influence the variable trend for intensity with a mean of 2.38. The mean scores for damage to individual and damage to environment were similar, 2.92 and 2.96, respectively. Out of a possible 28, the intensity total mean score was 12.38.

Duration. A review of the 4 items that are designed to measure the duration of problem behaviour showed the most restricted range of all the severity dimensions (5 out of a possible 7)

for 4 questions. No questions included the full range of responses for the duration dimension. The means for these questions is relatively stable across how soon after the behaviour reoccurs, how long it last, and for how long it takes for the behaviour to escalate and deescalate. The mean duration total was 15.36 out of a possible 28.

Discrimination. A review of the 4 items that are designed to measure the discrimination of behaviour revealed that caregivers reported a full range of responses from 1-7 for all questions except for when the behaviour occurs (6 out of a possible 7 responses). The means for duration questions showed a very stable trend across questions that assessed whom, across which setting, within which setting, and when the target behaviour occurs. The means for all questions were around 4 and the total discrimination mean was 18.28 out of a possible 28.

Caregivers' total severity scores. Mean caregiver's responses across dimension were most similar for frequency and discrimination responses; while intensity and duration were more variable mean scores. The caregivers' mean total severity score was 64.52 and ranged from 38 to 86 out of a possible 112. Compared to behaviour consultants (39-84), caregivers showed a wider range in overall responses of the TBSS.

Internal Consistency

An important component of the reliability of a scale is the internal consistency or degree to which the items on a particular scale relate to each other. Internal consistency measures are the "most widely used indices of reliability because they are the only ones that can be derived with only one administration of the test" (Streiner, Norman, & Cairney, 2014, p. 88). Cronbach's alpha coefficient was used to establish internal consistency of the total severity score and the dimension scores on the TBSS. Nunally (1978) was used for interpretation of the Cronbach's alpha. Particularly for pilot studies or early research, an alpha of .7 is acceptable on a measure

(Nunally, 1978). Both behaviour consultant ($\alpha = .757$) and caregiver ($\alpha = .711$) total severity score were found to have acceptable internal consistency (see Table 3). Internal consistencies for the dimension scales are depicted in Table 3. Internal consistency was calculated separately for the behaviour consultant's and the caregivers' as it was important to establish reliability separately for different user groups.

Table 3

Internal Consistency of the TBSS Dimension Totals

Dimension	Consultant Cronbach's alpha	Caregiver Cronbach's alpha
Frequency	.779	.549
Intensity	.913	.863
Duration	.570	.045
Discrimination	.733	.856
Total Severity	.757	.711

Note. Adequate Cronbach's alphas are boldface.

Consultants. The internal consistency of the behaviour consultant's ratings on the dimension scales was found over .90 for intensity, over .70 for frequency, .70 for discrimination and less than .60 for duration, which is unacceptable. Reliability analyses completed with this sample are reviewed and the most salient suggested changes to increase internal consistency are described to provide contextual information. Arguably the "most important" (Gliem & Gliem, 2003, p. 86) aspect of reliability, the "alpha if item deleted" statistic provides information separating the item from being questionable and acceptable. For the frequency scale, the removal of the question 1 (how often does this target behaviour occur) can increase internal consistency from $\alpha = .779$ to $\alpha = .887$. This may in fact, be too high, in that near perfect internal consistency can mean that some questions are redundant and not necessary (Cronbach, 1951). This is even more evident in the intensity dimension where the internal consistency is $\alpha = .913$ but could be decreased to eliminate redundancy to adequate internal reliability if either question 13, 14, or 15

were deleted. For the discrimination scale, the removal of question 9 (when does this target behaviour occur) can marginally increase the internal consistency from $\alpha = .733$ to $\alpha = .780$, both of which are considered acceptable. For the duration scale, if question 5 (how long is the de-escalation following the target behaviour) is deleted, the internal consistency can be decreased from $\alpha = .570$ to $\alpha = .247$, which is too low.

Caregivers. Internal consistency for the caregivers on the dimension scales above .80 for intensity and discrimination, and below .70 for frequency and duration. For intensity, if question 13 (physical damage to the person) was removed, internal consistency would decrease from $\alpha = .863$ to $\alpha = .390$. If question 9 (when does this behaviour occur) was removed under the discrimination dimension, internal consistency would increase from $\alpha = .856$ to $\alpha = 1$, an excellent rating. If question 1 (how often does this target behaviour occur) was removed under the frequency dimension, internal consistency would increase from $\alpha = .549$ to $\alpha = .702$, an acceptable consistency rating. For duration, internal consistency would increase from $\alpha = .045$ to $\alpha = .420$, if question 2 (how soon does the target behaviour reoccurs) was deleted.

Interrater Reliability

Mean comparison of consultant's and caregiver's ratings on the TBSS. The means of behaviour consultant's and caregiver's ratings on the 4 dimension scales and the total severity scores were completed using a paired sample *t*-test. Morgan, Leech, Gloeckner, and Barrett (2012) suggested that using a paired sample *t*-test for determining inter-rater reliability is preferred because it provides the comparison between the two raters as well as the correlations between their scores. The mean differences between behaviour consultant's and caregiver's responses are shown in Table 2 while the relationship between the mean differences are found in Table 4. There were significant differences in the mean scores on the discrimination dimension

as well as on the total severity score. In both cases the behaviour consultant's scores were found to be significantly lower than the caregiver scores. Though other scores followed a similar trend, the magnitude of the differences was not significant.

Table 4

Consultant's and Caregiver's TBSS Ratings Across Dimensions

Dimension	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Frequency	-1.040	4.954	-1.050	.304
Intensity	-.137	2.685	-.255	.801
Duration	-.320	1.701	-.941	.356
Discrimination	-2.760	6.307	-2.188	.039*
Total Severity	-4.257	10.249	-2.077	.049*

Note. * = $p < .05$

Relationship between consultant's and caregiver's ratings on the TBSS. The relationships between behaviour consultant's and caregiver's scores for each dimension were calculated as paired-sample correlations within the paired-sample *t*-test analysis in SPSS. An advantage of using this approach is that it limits the analyses to only the dimensions corresponding to the reliability analyses and does not consume statistical power on additional correlations. A summary of these correlations appears in Table 6.

Table 5

Cohen's 'Rule of Thumb' for Correlation Coefficient Interpretation

Size of the Correlation	Coefficient General Interpretation
.8 to 1.0	Very strong relationship
.6 to .8	Strong relationship
.4 to .6	Moderate relationship
.2 to .4	Weak relationships
.0 to .2	Weak or no relationship

The correlations were interpreted using Cohen's 'Rule of Thumb' found in Table 5. Very strong, significant, positive relationships were found between behaviour consultant and caregiver total scores on the intensity dimension and the duration dimension (see Table 6). There was a

strong significant ($p < .05$) positive relationship between behaviour consultant's and caregiver's total severity scores. There was a moderate significant positive relationship between behaviour consultant's and caregiver's frequency dimension total score. There was not a significant relationship between consultant's and caregiver's total scores on the discrimination dimension.

Table 6

Correlations between Consultant's and Caregiver's Responses

Consultant	Caregiver				TBSS Total
	Frequency	Intensity	Duration	Discrimination	
Frequency	.575**				
Intensity		.921**			
Duration			.912**		
Discrimination				.326	
TBSS Total					.693**

Note. * = $p < .05$; ** $p < .01$

Correlations of consultant TBSS total severity scores and the dimensions are shown in Table 7 and caregiver correlations between total dimensions of the TBSS are shown in Table 8. For consultants and caregivers, two significant but moderate relationships were found. There was a moderate significant positive relationship between duration and intensity for consultants and caregivers. Similarly, discrimination and frequency had a moderate significant positive relationship. As expected, the TBSS total severity scores were significantly related to each of the dimensions.

Table 7

Correlations Between Total Dimensions of TBSS from Consultants

Dimension	Frequency	Intensity	Duration	Discrimination	Total Severity
Frequency	-	-.291	-.113	.486*	.439*
Intensity		-	.522**	.28	.668**
Duration			-	.011	.545**
Discrimination				-	.746**
TBSS Total					-

Note. * = $p < .05$; ** $p < .01$

Table 8

Correlations Between Total Dimensions of TBSS from Caregivers

Dimension	Frequency	Intensity	Duration	Discrimination	Total Severity
Frequency	-	-.105	.121	.540**	.589**
Intensity		-	.510**	.036	.625**
Duration			-	.022	.603**
Discrimination				-	.667**
TBSS Total					-

Note. * = $p < .05$; ** $p < .01$

Face Validity

After examining the reliability of the TBSS, the next step was to examine face validity. Participants were given a series of questions that followed the 16 items such as “Does this tool measure how severe problem behaviour is?” Participants could respond by ‘yes’, ‘sort of’, or ‘no’. The 6 consultants who completed this question all responded favourably that the TBSS is an accurate measurement of the severity of problem behaviour. Eleven caregivers reported that the TBSS was an accurate tool to measure the severity of problem behaviour. Two caregivers disagreed while nine reported that it ‘sort of’ measured the severity. Two caregivers did not respond to this question.

Usability

In order to determine the usability of the measure, consultants and caregivers were instructed to indicate parts of the measure that were difficult to understand and then separately asked if the measure was easy to complete.

Hard to understand. Table 9 depicts the number of questions that were hard to understand for consultants and caregivers. The total questions marked are divided into dimensions; the ‘*n*’ depicts how many participants indicated a difficult question within that dimension. The totals at the bottom of the table are the amount of participants and questions marked as hard to understand across all measures.

Table 9

Hard to Understand Questions for Consultants and Caregivers

	Consultants (<i>n</i> = 6)		Caregivers (<i>n</i> = 25)	
Dimension	Total Questions Marked		Total Questions Marked	
Frequency	0	0	3	2
Intensity	0	0	3	3
Duration	1	2	1	2
Discrimination	1	1	3	3

Consultants. Two consultants indicated that three questions were difficult to understand.

One consultant found question 2 (duration-how soon does the target behaviour re-occur) difficult to understand, and another consultant found question 2 (duration- see above) and question 8 (discrimination-where does this target behaviour occur within settings) hard to understand.

Caregivers. Seven caregivers reported that they found some questions difficult to understand. In the frequency dimension, three questions were noted to be difficult to understand. Two caregivers found that question 10 (times of day) and question 11(times of month) difficult to understand. One caregiver reported that question 12 (time of year) was difficult to understand. In the intensity dimension, three questions were described as difficult to understand. Two caregivers noted that they found it difficult to understand question 13 (physical damage to the person). Another caregiver found question 14 (physical damage to others) difficult to understand. One caregiver identified questions 16 (disruption to routines) as difficult to understand. In the duration dimension, one question was noted to be difficult to understand. Two caregivers found question 2 (how soon does the target behaviour re-occur) hard to understand. In the discrimination dimension, caregivers noted three hard to understand questions. One caregiver found question 7 (occurs across settings) was difficult to understand. Another caregiver found question 8 (occurs within settings) hard to understand. Two caregivers indicated that question 9

(when i.e., specific activities/routines) was hard to understand. In summary, there were still relatively few concerns raised about the questions by caregivers.

Ease of completion. The TBSS included a close-ended question to assess the ease of completion for consultants and caregivers. Consultants reported that the TBSS was an easy tool to complete. Twenty-one percent of caregivers ($n = 5$) thought that the TBSS was ‘sort of’ easy to complete while 79% ($n = 19$) thought that it was easy to complete. Two caregivers did not respond.

Consultant and Caregiver Feedback on TBSS

Consultants and caregivers provided their impressions on how often they thought the TBSS should be completed. They were also asked open-ended questions regarding any improvements to the measure that were needed, or any other feedback that they wanted to share.

Frequency of completion. Consultants and caregivers were asked about the frequency of completing the TBSS. All consultants responded that it would be helpful to monitor client progress every 3 months ($n = 2$). Most caregivers responded that the TBSS should be used on a monthly basis (32%, $n = 8$) while others thought it should be used every 3 months (28%, $n = 7$). A few thought that every 6 months would be useful to monitor changes (16%, $n = 4$). Twelve percent thought that the TBSS should be conducted weekly and 1 person thought it should be used yearly (4%).

Improvements. In addition to the frequency of completing the measure, consultants and caregivers also asked for any additional feedback. Their specific responses will now be presented. The two consultants that provided feedback on the measure indicated that it was “easy to fill out” and “clear” as well as provided “depth” to the severity of problem behaviour. An idea to improve this question was to include the word ‘typically’ in the phrasing. One consultant

suggested that the anchors for “How often does this target behaviour occur?” be changed from ‘every minute’ to ‘every opportunity’. A final consultant suggestion was to “provide example for parents” within the questions to promote further understanding.

Caregivers. Caregivers provided more written feedback than consultants. Three parents indicated that the measure needed to be clearer because the questions were too ‘vague’. Another suggestion was to provide more choices in possible responses. One parent was unsure as to whether the measure was assessing a specific incident or the problem behaviour overall. There was some confusion over one question that asked, “How severe is the physical damage to the person as a result of the behaviour?” The respondent did not understand that the question was about the individual who displayed the problem behaviour. Other comments included leaving room for comments within each of the questions.

Discussion

ABA is an evidenced-based scientific practice that can be used to reduce problem behaviour and increase skills. Program evaluation of ABA services allows agencies to determine whether their specific outcomes are being met (e.g., to reduce problem behaviour and increase skills). However, specific challenges within the ABA model act as barriers to program evaluation. Individualized programming and evaluation presents a challenge to group evaluation designs, and a dearth of measures to capture discrete changes in problem behaviour over the course of assessment and treatment are some examples of the challenges that prevent program evaluation from occurring. In the present study, the reliability and validity of the TBSS was examined. Consultants and caregivers were both asked to complete a TBSS to rate the same target behaviour for a particular client. As a preliminary measure of the reliability of the scales, internal consistency was determined for the TBSS separately for consultants and caregivers. The

relationship of their scores was used to provide a preliminary indication of its inter-rater reliability. The mean scores of the consultants and caregivers were compared using paired sample *t*-tests and the relationship between participants' ratings was examined using correlations. To determine the face validity of the TBSS, consultants and caregivers were asked their impression of the degree to which the TBSS measured the severity of target behaviour. Additional information was collected pertaining to the how easy the TBSS was to complete, how often it should be completed and additional feedback. Overall the results of the study suggest that the TBSS is a promising measure that further efforts in refinement and development would be merited.

The results of the study will be organized and discussed in the context of the research questions that they were designed to answer; with discussion centered on the results and their implications for the future development of the TBSS with consideration of previous research findings, and relevant clinical applications and insights.

Research Question 1: Does the TBSS have acceptable internal consistency?

For both consultants and caregivers, internal consistency coefficients were acceptable for the total severity scores. Internal consistency was also acceptable for frequency, intensity, and discrimination dimension scores for consultants and for intensity and discrimination dimension scores for caregivers. Internal consistency was low for caregivers' frequency dimension score and unacceptable for duration dimension for behaviour consultants' and caregivers'.

These results are promising, but there is clear indication that some additional refinement to improve internal consistency would be helpful. One way to potentially increase the internal consistency is to increase the sample size (Gennarelli & Goodman, 2013). A second option could be to add more items on the measure, because fewer items on a measure are more likely to reach

absolute agreement due to fewer options (Stolarova et al., 2014). Therefore, increasing the total TBSS items from 16 to 20 may increase consistency across dimensions as well as between participants. Third, an increased sample size would reduce the chance of a Type II error (McCormick, Salcedo, & Poh, 2015). Preliminary results demonstrate acceptable internal consistency for consultants and caregivers on the total severity scale. Specific decisions to improve internal consistency will be considered in further analyses when a larger sample of data are available.

Research Question 2: What is the inter-rater reliability between consultants and caregivers on the TBSS?

There were no significant differences between mean consultant and caregiver total scores across frequency, intensity, and duration subscales (see Table 4); however, consultants' mean discrimination total score and mean total severity score were significantly lower than caregiver's scores.

There is limited research examining the factors influencing the difference between parent and behaviour consultant ratings on measures of problem behaviour, likely because there are limited measures of problem behaviour suitable for ABA-based interventions. Perhaps the difference between parent and consultant ratings may be accounted for by the consultant's knowledge and training in behaviour change, such as antecedent manipulations that can have an immediate and lasting impact on problem behaviour (e.g., Kern & Clemens, 2007). Another explanation might be that consultants may either be desensitized to the severity of problem behaviour or they may have a wider basis for comparison. Caregivers might provide higher severity ratings because they experience the impact of the problem behaviour first hand and for longer periods of time.

Overall, there were many positive correlations between consultant's and caregiver's responses of the TBSS. This result is consistent with other research where there are subtle differences between scores but overall similar agreement between clinician and parent ratings of therapy outcomes (i.e., Garcia, Joseph, Turk, & Basu, 2002; Rey, Plapp, & Simpson, 1999). The strongest relationships were between consultant and caregiver intensity and consultant and caregiver duration. With the exception of discrimination, every dimension of problem behaviour for consultants and caregivers had one or more variables with at least a moderate relationship, if not a very strong relationship. Very little relationship exists between the consultant's and caregiver's discrimination dimension. One possible explanation is that discrimination as a dimension of problem behaviour is less prevalent in the literature therefore parents may be less familiar with the descriptors whereas frequency and intensity are more common. Alternatively, parents may not see discrimination as a dimension related to problem behaviour or that the questions within the dimension are not related to one another (Tavakol & Dennick, 2011). Revisions can increase the internal consistency within discrimination. For example, it is possible that the term itself is uncommon, and perhaps a different term, such as pervasiveness or conditions would make the concept clearer to those completing the measure.

Rating discrepancies between caregiver and clinician are common in research (e.g., Bishop & Baird, 2001 or De Bildt et. al, 2004). Factors influencing the discrepancies may be based on the age and gender of individual or whether the problem behavior is internalizing or externalizing (Schroeder, Hood, & Hughes, 2010). Mental health of caregiver has also been shown to be a factor in rating discrepancies (Youngstrom, Loeber, & Stouthamer-Loeber, 2000; Berg-Neilson et al., 2003). Although an explanation as to why discrepancies exist between consultants and caregivers cannot be answered through the current study, the relevant finding is

that there is a difference between perspectives suggesting that each report is catching something that the other is not (Lemler, 2012). This highlights the sensitivity of the TBSS, which can be used to measure multiple perspective of the severity of behaviour. Future research should examine why consultants' and caregivers' scores were not more strongly positively correlated within the discrimination variable as well as why and what specific factors account for these differences.

Research Question 3: What are consultant's and caregiver's opinions of the face validity and usability of the TBSS?

Six consultants and 22 caregivers provided feedback on the face validity of the TBSS. All consultants ($n = 6$) and 50% of the caregivers ($n = 20$) agreed that the TBSS was an accurate measure of the severity of problem behaviour. Despite the fact some of the questions were reported to be difficult to understand for a few consultants and some caregivers, the TBSS has good internal consistency across total severity scores. Specifically, the intensity scale, completed by caregivers, had good internal consistency however intensity was the dimension with among the highest number of questions that were reported to be difficult to understand. These results demonstrate that some revisions to the intensity dimension may be in order.

Of consultants, 100% thought that the TBSS was easy to complete. Similarly, 79% of caregivers reported good usability for the TBSS, and 21% thought that it was 'sort of' easy to complete. Consultants reported that the TBSS should be used to monitor client progress every 3 months while only 28% of caregivers chose this frequency. The majority of caregivers thought that the TBSS should be used on a monthly basis (32%, $n = 8$), though other timeframes were suggested. Likely, the frequency of use should be decided by each agency, depending on the

schedule of their service provision. These results demonstrate good face validity and usability on the TBSS. Future research should examine larger sample sizes to validate these trends.

Consultants provided very positive feedback on the measure. They reported that the TBSS was clear and provided more depth into the measurement of problem behaviour. Caregivers, however, made some minor suggestions in terms of wording and specificity. These suggestions will be taken into account to improve the measure.

Limitations

One of the potential limitations was that the behaviour consultants recruited the caregivers to participate in the study. The physical presence of the consultant during the actual initialling of consent may have pressured the caregiver to participate in research. The consent process should prevent the participant from feeling obligated to participate in research. The consultant may have represented a position of power, thereby limiting free consent. To avoid this, consultants called caregivers using a phone script to obtain consent. Alternatively, involvement by a third-party researcher could complete caregiver recruitment to completely avoid coercion. Another limitation was the small sample size represented by the study. In addition, there was some missing data by two participants. However, the information that was missing accounted for less than 1% of the data, thereby having little effect over the actual results. Despite the smaller sample size, the study had several significant findings. The participant demographics were somewhat limited and further information regarding their overall education may affect their scores (i.e., do consultants with more education indicate more or less severe problem behaviour?). Finally, there were limited responses to the face validity and usability section of the TBSS. It is quite likely that participants did not choose to read the first question instructing participants to mark anywhere on the measure where it was difficult to understand.

As a result, there may have been unclear questions that were not indicated by participants. There were two questions that assessed the usability of the TBSS, which provided two methods of obtaining this information. One question instructed the participant to indicate if it was difficult to understand while another asked the participant if it was an easy measure to complete. Despite these limitations, the study shows promising results for the internal consistency of the total severity scale for both consultants and caregivers. Some changes should be made prior to use. Furthermore, the TBSS has been shown to reflect different perspective of consultants and caregiver, which can be used to demonstrate a robust representation of problem behaviour.

Future Research

The results of the current study suggest that it holds promise as a measure of the severity of problem behaviour. The internal consistency of the TBSS is consistent with other measures such as the DBC (O'Brien et al., 2001). Further testing against standardized measures should be conducted for further validation. The internal consistency of the measure, however, can still be improved. Therefore, additional testing of the measure should be conducted to obtain some data on test-retest reliability. Although there was relatively little feedback from the participants, future studies should address the feedback and improve the overall measure. The revised instrument should be piloted in the context of a clinical service so that multiple measurements can take place to examine the ability of the measure to detect change over time, which is its ultimate goal. Perspectives from caregivers and consultants demonstrated different feedback regarding the severity of problem behaviour. Direct observation is the preferred method of measurement in ABA (Kahng et al., 2001). Future research should compare the TBSS with direct observation to assess the accuracy of the measure. In addition, direct measures could be used to compare the accuracy of perspectives from consultants and caregivers. Due to the high cost of

direct observation (Campbell, 2002), the TBSS has potential to represent a viable alternative to the measurement of problem behaviour in certain circumstances (e.g. parent training groups) when objective data is not possible. Finally, problem behaviour has been shown to impact life quality (Lach et al., 2009; McIntyre, Blacher, & Baker, 2004; Renty & Roeyers, 2006), but there are few stand-alone measures that accurately measure problem behaviour (McClintock et al., 2003.). The next steps in this program of research will be to make refinements to the measure and pilot it in a prospective program evaluation to re-evaluate its reliability and validity, and examine its ability to distinguish change over time.

Conclusion

Developing a measure that can accurately represent various dimensions of problem behaviour is necessary in a step toward program evaluation within and across agencies that deliver behavioural interventions. The TBSS addresses some of the challenges within the ABA service model because the measure can be adapted to the individual's programming and can be used to track minor changes over time. The results in this study demonstrate that data from the consultant's and caregiver's responses provide varying perspectives that may enable a more holistic presentation of the severity of problem behaviour. The measure has acceptable internal consistency, good face validity, and good usability for consultants and caregivers. However, some revisions are in order to increase the internal consistency before the TBSS is ready for use among program evaluation. This is promising for individuals with DD because the measure can be used throughout assessment and treatment to assess changes in problem behaviour. The goal of the wider project is for agencies to demonstrate to funding sources that the desired outcomes are being met, therefore demonstrating the effective use of financial resources. Most importantly,

the individuals receiving services are meeting behavioural outcomes and their lives are being improved.

References

- Abbeduto, L., Seltzer, M. M., Shattuck, P., Krauss, M. W., Orsmond, G., & Murphy, M. M. (2004). Psychological well-being and coping in mothers of youths with autism, down syndrome, or fragile X syndrome. *American Journal on Mental Retardation*, 109(3). doi: 10.1111/j.1365-2788.2006.00907.x
- Achenbach, T. M. (1978). The child behavior profile: An empirically based system for assessing children's behavioral problems and competencies. *International Journal of Mental Health*, 24-42.
- Achenbach, T. M. (2009). *The Achenbach system of empirically based assessment (ASEBA): Development, findings, theory, and applications*. Burlington: University of Vermont Research Center for Children, Youth and Families.
- Achenbach, T. M., & Rescorla, L., A. (2000). *Manual for ASEBA Preschool Forms & Profiles*. Burlington, VT: Research Center for Children, Youth, and Families, University of Vermont.
- Achenbach, T., & Rescorla, L. (2013). Achenbach system of empirically based assessment. In *Encyclopedia of Autism Spectrum Disorders* (pp. 31-39). New York: Springer.
- Ahearn, W. H., Clark, K. M., MacDonald, R. P., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis*, 40(2), 263-275. doi: 10.1901/jaba.2007.30-06
- Alberta Learning. (2003). Managing challenging behaviour. In *Teaching students with autism spectrum disorders*. Retrieved from www.learning.gov.ab.ca/k_12/specialneeds

- Allison, J., Wilder, D. A., Chong, I., Lugo, A., Pike, J., & Rudy, N. (2012). A comparison of differential reinforcement and noncontingent reinforcement to treat food selectivity in a child with autism. *Journal of Applied Behavior Analysis*, 45(3), 613-617. doi: 10.1901/jaba.2012.45-613
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, D.C.: American Psychiatric Association.
- Athens, E. S., Vollmer, T. R., Sloman, K. N., & Pipkin, C. S. P. (2008). An analysis of vocal stereotypy and therapist fading. *Journal of Applied Behavior Analysis*, 41(2), 291. doi: 10.1901/jaba.2008.41-291.
- Atkinson, L., & Feldman, M. (1994). *Case study behaviour categories* (Unpublished measure). Toronto, ON.
- Azeem, M. W., Dogar, I. A., Shah, S., Cheema, M. A., Asmat, A., Akbar, M., ... & Haider, I. I. (2013). Anxiety and depression among parents of children with intellectual disability in pakistan. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 22(4), 290.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20(4), 313-327. doi: 10.1901/jaba.1987.20-313.
- Baio, J. (2014). Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. Morbidity and Mortality Weekly Report. *Surveillance Summaries*, 63(2), 1-21. Retrieved from <http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6302a1.htm>

- Baker, B. L., Blacher, J., Crnic, K. A., & Edelbrock, C. (2002). Behavior problems and parenting stress in families of three-year-old children with and without developmental delays. *Journal Information*, 107(6). doi: 10.1352/0895-8017(2002)107<0433:BPAPSI>2.0.CO;2
- Baker, B. L., McIntyre, L. L., Blacher, J., Crnic, K., Edelbrock, C., & Low, C. (2003). Pre-school children with and without developmental delay: Behaviour problems and parenting stress over time. *Journal of Intellectual Disability Research*, 47(4-5), 217-230. doi: 10.1046/j.1365-2788.2003.00484.x
- Bandyopadhyay, T. (2009). On measuring economic development. In S. Ghatak & P. Levine (Eds.), *Development macroeconomics: Essays in memory of Anita Ghatak* (pp. 49-67). New York, NY: Routledge.
- Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, 45(3), 601-613. doi: 10.1017/S003329171400172X
- Benson, P. R., & Karlof, K. L. (2009). Anger, stress proliferation, and depressed mood among parents of children with ASD: A longitudinal replication. *Journal of Autism and Developmental Disorders*, 39(2), 350-362. doi: 10.1007/s10803-008-0632-0
- Berg-Nielsen, T. S., Vika, A., & Dahl, A. A. (2003). When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health and Development*, 29(3), 207-213. doi: 10.1046/j.1365-2214.2003.00332.x
- Bicard, S. C, Bicard, D. F., & the IRIS Center. (2012). *Measuring behavior*. Retrieved from http://iris.peabody.vanderbilt.edu/wp-content/uploads/pdf_case_studies/ics_measbeh.pdf

- Bishop, D. V., & Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: Use of the children's communication checklist in a clinical setting. *Developmental Medicine and Child Neurology*, 43(12), 809-818. doi: 10.1017.S0012162201001475
- Bitsika, V., & Sharpley, C. F. (2004). Stress, anxiety and depression among parents of children with autism spectrum disorder. *Australian Journal of Guidance and Counselling*, 14(2), 151-161. doi: 10.1111/jftr.12015
- Blacher, J., & Baker, B. L. (2007). Positive impact of intellectual disability on families. *American Journal on Mental Retardation*, 112(5), 330-348.
- Bozkurt, F., & Gursel, O. (2005). Effectiveness of constant time delay on teaching snack and drink preparation skills to children with mental retardation. *Education and Training in Developmental Disabilities*, 390-400. Retrieved from <http://www.jstor.org/stable/23879956>
- Bowers, F. E., Jensen, M. E., Cook, C. R., McEachern, A. D., & Snyder, T. (2008). Improving the social status of peer-rejected youth with disabilities: Extending the research on positive peer reporting. *International Journal of Behavioral Consultation and Therapy*, 4(3), 230. doi: 10.1037/h0100851
- Brown, H. K., Ouellette-Kuntz, H., Bielska, I., & Elliott, D. (2009). Choosing a measure of support need: Implications for research and policy. *Journal of Intellectual Disability Research*, 53(11), 949-954. doi: 10.1111/j.1365-2788.2009.01216.x.

- Brown, M. A., McIntyre, L. L., Crnic, K. A., Baker, B. L., & Blacher, J. (2011). Preschool children with and without developmental delay: Risk, parenting, and child demandingness. *Journal of Mental Health Research in Intellectual Disabilities*, 4(3), 206-226. doi: 10.1080/19315864.2011.596990
- Buckles, J., Luckasson, R., & Keefe, E. (2013). A systematic review of the prevalence of psychiatric disorders in adults with intellectual disability, 2003–2010. *Journal of Mental Health Research in Intellectual Disabilities*, 6(3), 181-207. doi:10.1080/19315864.2011.651682
- Buescher, A. V., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA pediatrics*, 168(8), 721-728. doi: 10.1001/jamapediatrics.2014.210
- Burns, C. E. (2015). Does My Program Really Make a Difference? Program Evaluation Utilizing Aggregate Single-Subject Data. *American Journal of Evaluation*, 36(2), 191-203. doi: 10.1177/1098214014540032
- Cadman, D., Boyle, M., Szatmari, P., & Offord, D. R. (1987). Chronic illness, disability, and mental and social well-being: Findings of the Ontario child health study. *Pediatrics*, 79(5), 805-813. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2952939>
- Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: a quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 24(2), 120-138. doi: 10.1016/S0891-4222(03)00014-3
- Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis*, 18(2), 111-126.

- Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., McLaughlin, D. M., McAtee, M. L., ... & Braddock, D. E. (1999). Positive behavior support for people with developmental disabilities: A research synthesis. Washington, DC: American Association on Mental Retardation.
- Carr, E. G., Robinson, S., Taylor, J. C., & Carlson, J. I. (1990). Positive approaches to the treatment of severe behavior problems in persons with developmental disabilities: A review and analysis of reinforcement and stimulus-based procedures. Monograph No. 4. Seattle: The Association for Persons with Severe Handicaps.
- Centers for Disease Control and Prevention. (2010). Developmental disabilities. Retrieved from <http://www.cdc.gov/ncbddd/developmentaldisabilities/index.html>
- Cidav, Z., Marcus, S. C., & Mandell, D. S. (2012). Implications of childhood autism for parental employment and earnings. *Pediatrics*, 129(4), 617-623. doi: 10.1542/peds.2011-2700
- Cipani, E., & Schock, K. M. (2011). *Functional behavior assessment, diagnosis, and treatment: A complete system for education and mental health settings* (2nd ed.). New York, NY: Springer.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
- Condillac, R.A. (2009a). Evaluating processes and outcomes for behaviour therapy services: Phase 1- developing the Measures. Research Proposal Submitted to Brock University Research Ethics Board, St. Catharines, ON: Brock University.
- Condillac, R.A. (2009b). Impact on Quality of Life (IQOL). Unpublished Measure. St. Catharines, ON: Brock University.

Condillac, R.A. (2009c). The Behavioural Assessment and Treatment Taxonomy (BATT).

Unpublished Measure. St. Catharines, ON: Brock University.

Condillac, R.A. (2009d). The Target Behaviour Severity Scale (TBSS). Unpublished Measure.

St. Catharines, ON: Brock University.

Cooney, G., Jahoda, A., Gumley, A., & Knott, F. (2006). Young people with intellectual disabilities attending mainstream and segregated schooling: Perceived stigma, social comparison and future aspirations. *Journal of Intellectual Disability Research*, 50(6), 432-444. doi: 10.1111/j.1365-2788.2006.00789.x

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.).

Prentice Hall: Upper Saddle River, NJ.

Cooper, S. A., Smiley, E., Morrison, J., Williamson, A., & Allan, L. (2007). Mental ill-health in adults with intellectual disabilities Prevalence and associated factors. *The British Journal of Psychiatry*, 190(1), 27-35. doi: 10.1192/bjp.bp.106.022483

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship between autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, 34(2), 129-137. Retrieved from http://www.aspires-relationships.com/Interrelationship_Between_Autism_Diagnostic_Observation_Schedule_Generic.pdf

- De Bildt, A., Sytema, S., Kraijer, D., Sparrow, S., & Minderaa, R. (2005). Adaptive functioning and behaviour problems in relation to level of education in children and adolescents with intellectual disability. *Journal of Intellectual Disability Research*, 49(9), 672-681. doi: 10.1111/j.1365-2788.2005.00711.x
- Dekker, M. C., Nunn, R. J., Einfeld, S. E., Tonge, B. J., & Koot, H. M. (2002). Assessing emotional and behavioral problems in children with intellectual disability: Revisiting the factor structure of the Developmental Behavior Checklist. *Journal of Autism and Developmental Disorders*, 32(6), 601-610. doi: 10.1023/A:1021263216093
- Dib, N., & Sturmey, P. (2007). Reducing student stereotypy by improving teachers' implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, 40(2), 339-343. doi: 10.1901/jaba.2007.52-06
- Dickman, S. E., Bright, C. N., Montgomery, D. H., & Miguel, C. F. (2012). The effects of response interruption and redirection (RIRD) and differential reinforcement on vocal stereotypy and appropriate vocalizations. *Behavioral Interventions*, 27(4), 185-192. doi: 10.1002/bin.1348
- Didden, R., Sturmey, P., Sigafos, J. Lang, R., O'Reilly, M. F., Lancioni, G. E. (2012). Nature, prevalence, and characteristics of challenging behaviour. In J. L., Matson (Ed.), *Functional assessment for challenging behaviors* (pp. 25-44). New York: SpringerDoggett, R. A., Edwards, R. P., Moore, J. W., Tingstrom, D. H., & Wilczynski, S. M. (2001). An approach to functional assessment in general education classroom settings. *School Psychology Review*, 30(3), 313. doi: 10.1037/h0088301

- Donenberg G., & Baker B. L. (1993) The impact of young children with externalizing behaviors on their families. *Journal of Abnormal Child Psychology* 21(2), 179–98.
doi: 10.1007/BF00911315
- Durand, V. M. (1990). *Severe behavior problems: A functional communication training approach*. Guilford Press.
- Durand, V. M., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. *Journal of Applied Behavior Analysis*, 24(2), 251-264. doi: 10.1901/jaba.1991.24-251
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4-to 7-year-old children with autism a 1-year comparison controlled study. *Behavior Modification*, 26(1), 49-68. doi: 10.1177/0145445502026001004
- Einfeld S. L., & Tonge B. J. (2002). *Manual for the Developmental Behaviour Checklist: Primary Carer Version (DBC- P) and Teacher Version (DBC-T)*, 2nd ed. Monash University Centre for Developmental Psychiatry and Psychology, Melbourne.
- Einfeld, S. L., Ellis, L. A., & Emerson, E. (2011). Comorbidity of intellectual disability and mental disorder in children and adolescents: A systematic review. *Journal of Intellectual and Developmental Disability*, 36(2), 137-143. doi:10.1080/13668250.2011.572548
- Eisenhower, A. S., Baker, B. L., & Blacher, J. (2005). Preschool children with intellectual disability: Syndrome specificity, behaviour problems, and maternal well-being. *Journal of Intellectual Disability Research*, 49(9), 657-671. doi: 10.1111/j.1365-2788.2005.00699.x

- Emerson, E., & Hatton, C. (2007). Mental health of children and adolescents with intellectual disabilities in Britain. *The British Journal of Psychiatry*, *191*(6), 493-499.
doi: 10.1192/bjp.bp.107.038729
- Feldman, M. A., & Werner, S. E. (2002). Collateral effects of behavioral parent training on families of children with developmental disabilities and behavior disorders. *Behavioral Interventions*, *17*(2), 75-83. doi: 10.1002/bin.111
- Feldman, M. A., Atkinson, L., Foti-Gervais, L., & Condillac, R. (2004). Formal versus informal interventions for challenging behaviour in persons with intellectual disabilities. *Journal of Intellectual Disability Research*, *48*(1), 60-68. doi: 10.1111/j.1365-2788.2004.00578.x
- Feldman, M. A., Condillac, R. A., Tough, S. E., Hunt, S., & Griffiths, D. (2002). Effectiveness of community positive behavioral intervention for persons with developmental disabilities and severe behavioral challenges. *Behavior Therapy*, *33*(3), 377-398.
doi: 10.1016/S0005-7894(02)80034-X
- Feldman, M. A., Hancock, C. L., Rielly, N., Minnes, P., & Cairns, C. (2000). Behavior problems in young children with or at risk for developmental delay. *Journal of Child and Family Studies*, *9*(2), 247-261. doi: 10.1023/A:1009427306953
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.
- Freedman, B. H., Kalb, L. G., Zablotsky, B., & Stuart, E. A. (2012). Relationship status among parents of children with autism spectrum disorders: A population-based study. *Journal of Autism and Developmental Disorders*, *42*(4), 539-548. doi: 10.1007/s10803-011-1269-y

- Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, 45(2), 449-453. doi: 10.1901/jaba.2012.45-449
- Gaber, J. (2000). Meta-needs assessment. *Evaluation and Program Planning*, 23(2), 139-147. doi:10.1016/S0149-7189(00)00012-4
- Gambrill, E. (2012). Birds of a feather: Applied behavior analysis and quality of Life. *Research on Social Work Practice*, 23(2), 121-140. doi:10.1177/1049731512465775
- Ganz, M. L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics & Adolescent Medicine*, 161(4), 343-349. doi: 10.1001/archpedi.161.4.343
- Garcia, R., Joseph, T., Turk, J., & Basu, R. (2002). A comparison of parent and therapist ratings of outcome in a child mental health clinic. *Child and Adolescent Mental Health*, 7(4), 168-172. doi: 10.1111/1475-3588.00030
- Gennarelli, R., & Goodman, M. S. (2013). Measuring internal consistency of community engagement using the APLHA option of PROC CORR. Paper presented at the NorthEast SAS Users Group Burlington, VT. Retrieved from https://www.researchgate.net/profile/Melody_Goodman/publication/260782196_Measuring_Internal_Consistency_of_Community_Engagement_using_the_APLHA_option_of_PROC_CORR/links/02e7e53231f173255f000000.pdf
- Gliem, R. R., & Gliem, J. A. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales. Paper presented at the Midwest research-to-practice conference in adult, continuing, and community education, OH. Retrieved from <https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem%20&%20Gliem.pdf?s..>

- Gray, D. E. (1993). Perceptions of stigma: The parents of autistic children. *Sociology of Health & Illness*, 15(1), 102-120. doi: 10.1111/1467-9566.ep11343802
- Grey, I. M., Honan, R., McClean, B., & Daly, M. (2005). Evaluating the effectiveness of teacher training in applied behaviour analysis. *Journal of Intellectual Disabilities*, 9(3), 209-227. doi: 10.1177/1744629505056695
- Griffith, G. M., Hastings, R. P., Nash, S., & Hill, C. (2010). Using matched groups to explore child behavior problems and maternal well-being in children with Down syndrome and autism. *Journal of Autism and Developmental Disorders*, 40(5), 610-619. doi: 10.1007/s10803-009-0906-1
- Griffiths, D. M., & Watson, S. L. (2004). Demystifying syndromes associated with developmental disabilities. In D.M. Griffiths & R. King (Eds.), *Demystifying syndromes: Clinical and educational implications of common syndromes associated with persons with intellectual disabilities* (pp. 1-30). Kingston, NY: NADD.
- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis*, 36, 147–185. doi: 10.1901/jaba.2003.36-147
- Harris, J. C. (2006). *Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment*. New York: Oxford University Press.
- Hartley, S. L., Barker, E. T., Seltzer, M. M., Floyd, F., Greenberg, J., Orsmond, G., & Bolt, D. (2010). The relative risk and timing of divorce in families of children with an autism spectrum disorder. *Journal of Family Psychology*, 24(4), 449. doi: 10.1037/a0019847

- Hastings, R. P., & Brown, T. (2002). Behavior problems of children with autism, parental self-efficacy, and mental health. *Journal Information*, 107(3). Retrieved from <http://www.w.aspires-relationships.com/hastings6.pdf.pdf>
- Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. *Journal of Vocational Rehabilitation*, 32(2), 125.
doi:10.3233/JVR-2010-0502
- Herring, S., Gray, K., Taffe, J., Tonge, B., Sweeney, D., & Einfeld, S. (2006). Behaviour and emotional problems in toddlers with pervasive developmental disorders and developmental delay: associations with parental mental health and family functioning. *Journal of Intellectual Disability Research*, 50(12), 874-882. doi: 10.1111/j.1365-2788.2006.00904.x
- Heyvaert, M., Saenen, L., Campbell, J. M., Maes, B., & Onghena, P. (2014). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: An updated quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, 35(10), 2463-2476. doi: 10.1016/j.ridd.2014.06.017
- Hill, J., & Furniss, F. (2006). Patterns of emotional and behavioural disturbance associated with autistic traits in young people with severe intellectual disabilities and challenging behaviours. *Research in Developmental Disabilities*, 27(5), 517-528.
doi:10.1016/j.ridd.2005.07.001
- Horlin, C., Falkmer, M., Parsons, R., Albrecht, M. A., and Falkmer, T. (2014). The cost of autism spectrum disorders. *PLoS ONE* 9(9): e106552. doi: 10.1371/journal.pone.0106552

- Horner, R. H., Carr, E. G., Strain, P. S., Todd, A. W., & Reed, H. K. (2002). Problem behavior interventions for young children with autism: A research synthesis. *Journal of Autism and Developmental Disorders*, 32(5), 423-446. doi: 10.1023/A:1020593922901
- Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, 26(4), 359-383. doi:10.1016/j.ridd.2004.09.005
- Hughes, C., Fowler, S. E., Copeland, S. R., Agran, M., Wehmeyer, M. L., & Church-Pupke, P. P. (2004). Supporting high school students to engage in recreational activities with peers. *Behavior modification*, 28(1), 3-27. doi: 10.1177/0145445503259215
- Hurwitz, S., & Minshawi, N. F. (2012). Methods of defining and observing behaviors. In J. L., Matson (Ed.), *Functional assessment for challenging behaviours* (pp. 91-103). New York: Springer.
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, 1(1), 3. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2846577/>
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, 27(2), 197. doi: 10.1901/jaba.1994.27-197
- Kahng, S., Ingvarsson, E. T., Quigg, A. M., Seckinger, K. E., & Teichman, H. M. (2011). Defining and measuring behavior. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of Applied Behaviour Analysis* (pp. 113-131). New York: Guilford Press.

- Kaiser, A. P., Hancock, T. B., & Nietfeld, J. P. (2000). The effects of parent-implemented enhanced milieu teaching on the social communication of children who have autism. *Early Education and Development, 11*(4), 423-446. doi: 10.1207/s15566935eed1104_4
- Kanner L. (1943). Autistic disturbances of affective contact. *Nervous Child 2*, 217-250.
- Kats, D., Payne, L., Parlier, M., & Piven, J. (2013). Prevalence of selected clinical problems in older adults with autism and intellectual disability. *Journal of Neurodevelopmental Disorders, 5*(1), 27. doi:10.1186/1866-1955-5-27
- Kazdin, A. E., Siegel, T. C., & Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology, 60*(5), 733. Retrieved from <http://dx.doi.org/10.1037/0022-006X.60.5.733>
- Kelley, M. E., LaRue, R. H., Roane, H. S., & Gadaire, D. M. (2011). Indirect behavioral assessments. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of Applied Behaviour Analysis* (pp. 182-190). New York: Guilford Press.
- Kennedy, C. H. (2005). *Single case designs for educational research*. Boston: Pearson.
- Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools, 44*(1), 65. doi:10.1002/pits.20206
- Knapp, M., Comas-Herrera, A., Astin, J., Beecham, J., & Pendaries, C. (2005). Intellectual disability, challenging behaviour and cost in care accommodation: What are the links?. *Health & Social Care in the Community, 13*(4), 297-306. <http://dx.doi.org/10.1111/j.1365-2524.2005.00539.x>

- Koegel, L. K., Koegel, R. L., & Dunlap, G. (1996). *Positive behavioral support: Including people with difficult behavior in the community*. Baltimore, MD: Paul H. Brookes Publishing.
- Koegel, R. L., & Koegel, L. K. (2006). *Pivotal response treatments for autism: Communication, social, and academic development*. Baltimore, MD: Brookes Publishing.
- Koskentausta, T., & Almqvist, F. (2004). Developmental Behaviour Checklist (DBC) in the assessment of psychopathology in Finnish children with intellectual disability. *Journal of Intellectual and Developmental Disability*, 29(1), 27-39. doi: 10.1080/13668250410001662883
- Kratochwill, T. R., & Van Someren, K. R. (1995). Barriers to treatment success in behavioral consultation: Current limitations and future directions. *Journal of Educational and Psychological Consultation*, 6(2), 125-143. doi: 10.1207/s1532768xjepc0602_3
- Kupzyk, S., Daly, E. J., & Andersen, M. N. (2011). A comparison of two flash-card methods for improving sight-word reading. *Journal of Applied Behavior Analysis*, 44(4), 781-792. Retrieved from <http://doi.org/10.1901/jaba.2011.44-781>
- Kurtz, P. F., Chin, M. D., Huete, J. M., Tarbox, R. S., O'Connor, J. T., Paclawskyj, T. R., & Rush, K. S. (2003). Functional analysis and treatment of self-injurious behavior in young children: A summary of 30 cases. *Journal of Applied Behavior Analysis*, 36(2), 205-219. doi: 10.1901/jaba.2003.36-205
- Lach, L. M., Kohen, D. E., Garner, R. E., Brehaut, J. C., Miller, A. R., Klassen, A. F., & Rosenbaum, P. L. (2009). The health and psychosocial functioning of caregivers of children with neurodevelopmental /disorders. *Disability and Rehabilitation*, 31(8), 741-752. doi: 10.1080/09638280802242163

- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., & Sigafoos, J. (2009). An overview of behavioral strategies for reducing hand-related stereotypies of persons with severe to profound intellectual and multiple disabilities: 1995–2007. *Research in Developmental Disabilities*, 30(1), 20-43. doi: 10.1016/j.ridd.2008.02.002
- Lanovaz, M. J., & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification*, 36(2), 146-164. doi: 10.1177/0145445511427192
- Lauritsen, M. B. (2013). Autism spectrum disorders. *European Child & Adolescent Psychiatry*, 22(1), 37-42. doi: 10.1007/s00787-012-0359-5
- LaVigna, G. W., & Willis, T. J. (2012). The efficacy of positive behavioural support with the most challenging behaviour: The evidence and its implications. *Journal of Intellectual and Developmental Disability*, 37(3), 185-195. doi: 10.3109/13668250.2012.696597
- Lecavalier, L. (2006). Behavioral and emotional problems in young people with pervasive developmental disorders: Relative prevalence, effects of subject characteristics, and empirical classification. *Journal of Autism and Developmental Disorders*, 36(8), 1101-1114. doi: 10.1007/s10803-006-0147-5
- Lecavalier, L., Leone, S., & Wiltz, J. (2006). The impact of behaviour problems on caregiver stress in young people with autism spectrum disorders. *Journal of Intellectual Disability Research*, 50(3), 172-183. Retrieved from <http://dx.doi.org/10.1111/j.1365-2788.2005.00732.x>
- Lee, Y. Y., Sugai, G., & Horner, R. H. (1999). Using an instructional intervention to reduce problem and off-task behaviors. *Journal of Positive Behavior Interventions*, 1(4), 195-204. doi: 10.1177/109830079900100402

- Lemler, M. (2012). Discrepancy between parent report and clinician observation of symptoms in children with autism spectrum disorders. *Discussions*, 8(2). Retrieved from <http://www.studentpulse.com/articles/803/3/discrepancy-between-parent-report-and-clinician-observation-of-symptoms-in-children-with-autism-spectrum-disorders>
- Lewis, T. J., Scott, T. M., & Sugai, G. (1994). The problem behavior questionnaire: A teacher-based instrument to develop functional hypotheses of problem behavior in general education classrooms. *Assessment for Effective Intervention*, 19(2-3), 103-115.
doi: 10.1177/073724779401900207
- Lowe K., Felce D., & Blackman D. (1995) People with learning disabilities and challenging behaviour: the characteristics of those referred and not referred to specialist teams. *Psychological Medicine* 25(3), 595–603. doi: 10.1017/S003329170003350X
- MacDonald, R., Green, G., Mansfield, R., Geckeler, A., Gardenier, N., Anderson, J., ... & Sanchez, J. (2007). Stereotypy in young children with autism and typically developing children. *Research in Developmental Disabilities*, 28(3), 266-277.
- Maenner, M. J., Rice, C. E., Arneson, C. L., Cunniff, C., Schieve, L. A., Carpenter, L. A., ... & Durkin, M. S. (2014). Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. *JAMA psychiatry*, 71(3), 292-300.
doi: 10.1001/jamapsychiatry.2013.3893
- Mannion, A., Leader, G., & Healy, O. (2013). An investigation of comorbid psychological disorders, sleep problems, gastrointestinal symptoms and epilepsy in children and adolescents with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 7(1), 35-42. doi: 10.1016/j.rasd.2012.05.002

- Martinez, C. K., & Betz, A. M. (2013). Response interruption and redirection: Current research trends and clinical application. *Journal of Applied Behavior Analysis*, 46(2), 549-554. doi: 10.1002/jaba.38
- Martorell, A., Gutierrez-Recacha, P., Pereda, A., & Ayuso-Mateos, J. L. (2008). Identification of personal factors that determine work outcome for adults with intellectual disability. *Journal of Intellectual Disability Research*, 52(12), 1091-1101. doi:10.1111/j.1365-2788.2008.01098.x
- Matson, J. L., & Nebel-Schwalm, M. S. (2007). Comorbid psychopathology with autism spectrum disorder in children: An overview. *Research in Developmental Disabilities*, 28(4), 341-352. doi: 10.1007/s40489-014-0012-y
- McClintock, K., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: A meta-analytic study. *Journal of Intellectual Disability Research*, 47(6), 405-416. doi: 10.1046/j.1365-2788.2003.00517.x
- McCormick, K., Salcedo, J., & Poh, A. (2015). *SPSS Statistics for Dummies* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- McDonough, J. T., & Revell, G. (2010). Accessing employment supports in the adult system for transitioning youth with autism spectrum disorders. *Journal of Vocational Rehabilitation*, 32(2), 89. DOI: 10.3233/JVR-2010-0498
- McGill, P., Emerson, E., & Mansell, J. (1994). Individually designed residential provision for people with seriously challenging behaviours. In *Severe learning disabilities and challenging behaviours* (pp. 119-156). United Kingdom: Springer. doi: 10.1007/978-1-4899-2961-7

- McIntyre, L. L., Blacher, J., & Baker, B. L. (2002). Behaviour/mental health problems in young adults with intellectual disability: The impact on families. *Journal of Intellectual Disability Research*, 46(3), 239-249. doi: 10.1046/j.1365-2788.2002.00371.x
- Miltenberger, R. (2008). *Behaviour Modification*. Belmont, CA. Wadsworth Publishing.
- Mead, D. E. (2002). Marital distress, co-occurring depression, and marital therapy: A review. *Journal of Marital and Family Therapy*, 28(3), 299-314. doi: 10.1111/j.1752-0606.2002.tb01188.x
- Moes, D. (1995). Parent education and parenting stress. In R. L. Koegel & L. K. Koegel (Eds.), *Teaching children with autism: Strategies for initiating positive interactions and improving learning opportunities* (pp. 79–93). Baltimore, MD: Brookes Publishing.
- Moore, T. R., & Symons, F. J. (2011). Adherence to treatment in a behavioral intervention curriculum for parents of children with autism spectrum disorder. *Behavior Modification*, 35(6), 570-594. doi: 10.1177/0145445511418103
- Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2012). *IBM SPSS for introductory statistics: Use and interpretation* (pp. 116-117). Retrieved from https://tandfbis.s3.amazonaws.com/rt-media/pdf/9781848729827/pages116_117.pdf
- Moss, S., Prosser, H., Costello, H., & Simpson, N. (1998). Reliability and validity of the PAS-ADD Checklist for detecting psychiatric disorders in adults with intellectual disability. *Journal of Intellectual Disability Research*, 42(2), 173-183. doi: 10.1046/j.1365-2788.1998.00116.x
- Napolitano, D. A., Madaus Knapp, V., Speares, E., McAdam, D. B., & Brown, H. (2012). The role of functional assessment in treatment planning. In J. L., Matson (Ed.), *Functional assessment for challenging behaviors* (pp. 195-212). New York: Springer.

National Research Council (2001). *Educating children with autism*. National Academy Press.

Washington, DC.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Brien, G., Pearson, J., Berney, T., & Barnard, L. (2001). Measuring behaviour in

developmental disability: A review of existing schedules. *Developmental Medicine and Child Neurology* 43(87), 1-72. Retrieved from

<http://onlinelibrary.wiley.com/store/10.1111/j.1469-8749.2001.tb12454.x/asset/j.1469-8749.2001.tb12454.x.pdf?v=1&t=id7dkg35&s=c8bdf35d89244c765ed1fbc7b8289ff054ec73bb>

Olsson, M. B., & Hwang, C. P. (2001). Depression in mothers and fathers of children with intellectual disability. *Journal of Intellectual Disability Research*, 45(6), 535-543.

doi: 10.1046/j.1365-2788.2001.00372.x

Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice*, 5(2), 2. Retrieved from

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3592486/>

Patchin, J. W., & Hinduja, S. (2010). Cyberbullying and self-esteem. *Journal of School Health*, 80(12), 614-621. doi: 10.1111/j.1746-1561.2010.00548.x

Perry, A. (2004). A model of stress in families of children with developmental disabilities:

Clinical and research applications. *Journal on Developmental Disabilities*, 11(1), 1-16.

Pipkin, C. S. P., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis*, 43(1), 47-70. doi: 10.1901/jaba.2010.43-47

Posavac, E. J., & Carey, R. G. (2007). *Program evaluation: Methods and case studies (7th ed.)*.

Upper Saddle River, NJ: Pearson Education, Inc.

Reid, D. H., Parsons, M. B., & Green, C. W. (2012). *The supervisor's guidebook: Evidence-based strategies for promoting work quality and enjoyment among human service staff*.

Morgantown, NC: Habilitative Management Consultants, Inc.

Reid, D.H., O'Kane, N.P., and Macurik, K.M. (2011). Staff training and management. In W.W.

Fisher, C. Piazza, & H. Roane (Eds). *Handbook of Applied Behaviour Analysis* (p. 281-294). NY: Guilford Publications.

Reiss, S., & Rojahn, J. (1993). Joint occurrence of depression and aggression in children and adults with mental retardation. *Journal of Intellectual Disability Research*, 37(3), 287-294. doi: 10.1111/j.1365-2788.1993.tb01285.x

Renty, J. O., & Roeyers, H. (2006). Quality of life in high-functioning adults with autism spectrum disorder: The predictive value of disability and support characteristics. *Autism*, 10(5), 511-524. doi: 10.1177/1362361306066604

Rey, J. M., Plapp, J. M., & Simpson, P. L. (1999). Parental satisfaction and outcome: A 4-year study in a child and adolescent mental health service. *Australian and New Zealand Journal of Psychiatry*, 33(1), 22-28. doi:10.1046/j.1440-1614.1999.00516.x

Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2001). The Behavior Problems Inventory: An instrument for the assessment of self-injury, stereotyped behavior, and aggression/destruction in individuals with developmental disabilities. *Journal of Autism and Developmental Disorders*, 31(6), 577-588. doi: 10.1023/A:1013299028321

- Rojahn, J., Rowe, E. W., Sharber, A. C., Hastings, R., Matson, J. L., Didden, R., ... & Dumont, E. L. M. (2012). The Behavior Problems Inventory-Short Form for individuals with intellectual disabilities: Part II: reliability and validity. *Journal of Intellectual Disability Research*, 56(5), 546-565. doi: 10.1111/j.1365-2788.2011.01507.x
- Rojahn, J., Schroeder, S. R., Mayo-Ortega, L., Oyama-Ganiko, R., LeBlanc, J., Marquis, J., & Berke, E. (2013). Validity and reliability of the Behavior Problems Inventory, the Aberrant Behavior Checklist, and the Repetitive Behavior Scale-Revised among infants and toddlers at risk for intellectual or developmental disabilities: A multi-method assessment approach. *Research in Developmental Disabilities*, 34(5), 1804-1814. doi: 10.1016/j.ridd.2013.02.024
- Rojahn, J., Wilkins, J., Matson, J. L., & Boisjoli, J. (2010). A comparison of adults with intellectual disabilities with and without ASD on parallel measures of challenging behaviour: The Behavior Problems Inventory-01 (BPI-01) and Autism Spectrum Disorders-Behavior Problems for intellectually disabled adults (ASD-BPA). *Journal of Applied Research in Intellectual Disabilities*, 23(2), 179-185. doi: 10.1111/j.1468-3148.2009.00519.x
- Schalock, R. L., & Verdugo, M. Á. (2012). *A leadership guide for today's disabilities organizations: Overcoming challenges and making change happen*. Baltimore: Paul H. Brookes Publishing Co.
- Schreibman, L., Heyser, L., & Stahmer, A. (1999). Autistic disorder: Characteristics and behavioral treatment. In N. A., Wieseler & R. H., Hanson (Eds.), *Challenging behavior of persons with mental health disorders and severe disabilities* (pp. 39-63). Washington, DC: American Association of Mental Retardation.

- Schroeder, J. F., Hood, M. M., & Hughes, H. M. (2010). Inter-parent agreement on the syndrome scales of the Child Behavior Checklist (CBCL): Correspondence and discrepancies. *Journal of Child and Family Studies*, 19(5), 646-653. doi: 10.1007/s10826-010-9352-0
- Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state-of-the-art. *Journal of Applied Behavior Analysis*, 24(2), 189-204. doi: 10.1901/jaba.1991.24-189
- Services and Supports to Promote the Social Inclusion of Persons with Developmental Disabilities Act, S.O. c.14. (2008). Retrieved from <http://www.ontario.ca/laws/statute/08s14#BK0>
- Shtayermman, O. (2013). Stress and marital satisfaction of parents to children diagnosed with autism. *Journal of Family Social Work*, 16(3), 243-259. doi: 10.1080/10522158.2013.786777
- Sigafoos, J., Arthur, M., & O'Reilly, M. F. (2003). *Challenging behaviour and developmental disability*. London: Whurr Publishers.
- Smith, M. D., & Philippen, L. R. (2005). Community integration and supported employment. In D. Zager (Ed.), *Autism spectrum disorders: Identification, education, and treatment* (3rd ed., pp. 493-514). New Jersey: Lawrence Erlbaum Associates, Inc.
- Statistics Canada. (2013). Family caregiving: What are the consequences? Statistics Canada Catalogue no. 75-006-X. Retrieved from <http://www.statcan.gc.ca/pub/75-006-x/2013001/article/11858-eng.pdf>
- Steege, M. W., Mace, F. C., Perry, L., & Longenecker, H. (2007). Applied behavior analysis: Beyond discrete trial teaching. *Psychology in the Schools*, 44(1), 91-99. doi: 10.1002/pits.20208

- Stolarova, M., Wolf, C., Rinker, T., & Brielmann, A. (2014). How to assess and compare interrater reliability, agreement and correlation of ratings: An exemplary analysis of mother-father and parent-teacher expressive vocabulary rating pairs. *Frontiers in Psychology*, 5(509), 1-13. doi: 10.3389/fpsyg.2014.00509
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., ... & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions*, 2(3), 131-143. doi: 10.1177/109830070000200302
- Tabachnick, B. G., & Fidell, L. S. (2001). Using multivariate statistics (4th ed.). Boston: Allyn & Bacon.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. doi: 10.5116/ijme.4dfb.8dfd s
- Thompson, R. H., & Borrero, J. C. (2011). Direct observation. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of Applied Behaviour Analysis* (pp. 191-205). New York: Guilford Press.
- Tonge, B. J., & Brereton, A. V. (2011). Overview of the developmental behaviour checklist. Sydney, Australia: Monash University. Retrieved from <http://www.med.monash.edu.au/assets/docs/scs/psychiatry/developmental-behaviour-checklist.pdf>
- Tureck, K., Matson, J. L., & Beighley, J. S. (2013). An investigation of self-injurious behaviors in adults with severe intellectual disabilities. *Research in Developmental Disabilities*, 34(9), 2469-2474. doi: 10.1016/j.ridd.2013.05.022

- Turnbull, A. P., & Ruef, M. (1997). Family perspectives on inclusive lifestyle issues for people with problem behavior. *Exceptional Children*, 63(2), 211-227. doi: 10.1177/001440299706300205
- Vollmer, T. R. & Athens, E. (2011). Developing function-based extinction procedures for problem behaviour. In W. Fisher, C. Piazza, & H. Roane (Eds.). *Handbook of Applied Behaviour Analysis* (pp. 317-335). New York, NY: The Guilford Press.
- Walsh, C. E., Mulder, E., & Tudor, M. E. (2013). Predictors of parent stress in a sample of children with ASD: Pain, problem behavior, and parental coping. *Research in Autism Spectrum Disorders*, 7(2), 256-264. doi:10.1016/j.rasd.2012.08.010
- Weitlauf A. S., McPheeters M. L., Peters B., Sathe N., Travis R., Aiello R., Williamson E., Veenstra-VanderWeele J., Krishnaswami S., Jerome R., & Warren Z. (2014). Therapies for children with autism spectrum disorder: Behavioral interventions update [Internet]. Comparative Effectiveness Review No. 137. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK241444/>
- Wilczynski, S. M. (2011). In J. K. Luiselli (Ed.), *Teaching and behavior support for children and adults with autism spectrum disorder: A practitioner's guide* (pp. 55-62). New York: Oxford University Press.
- Winn S., & Hay I. (2009). Transition from school for youths with a disability: Issues and challenges. *Disability and Society* 24(1), 103–115. doi: 10.1177/1088357608329827
- Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203–214. doi: 10.1901/jaba.1978.11-203

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, 68(6), 1038. doi: 10.1037//0022-006X.68.6.1038